AGARD-CP-575

AGARD-CP-575

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 575

Situation Awareness: Limitations and Enhancement in the Aviation Environment

(la Conscience de la situation: les limitations et l'amélioration en environnement aéronautique)

Papers presented at the Aerospace Medical Panel Symposium held in Brussels, Belgium from 24-27 April 1995.

NORTH ATLANTIC TREATY ORGANIZATION

Published January 1996

Distribution and Availability on Back Cover

AGARD-CP-575

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 575

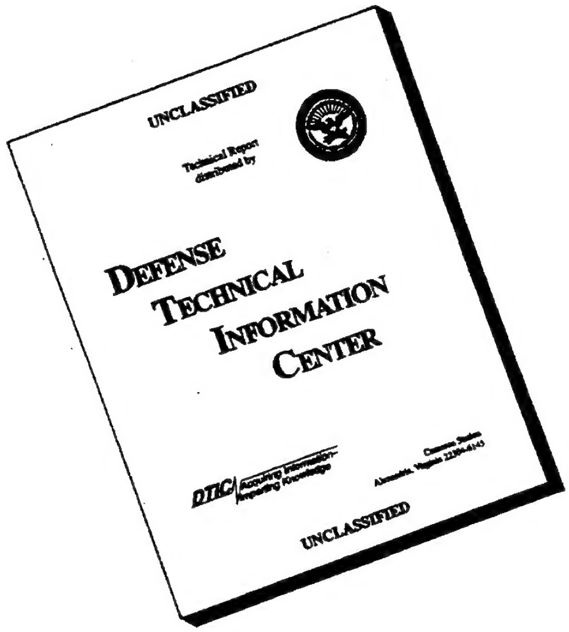
Situation Awareness: Limitations and Enhancement in the Aviation Environment

(la Conscience de la situation: les limitations et l'amélioration en environnement aéronautique)

Papers presented at the Aerospace Medical Panel Symposium held in Brussels, Belgium from 24-27 April 1995.

North Atlantic Treaty Organization
Organisation du Traité de l'Atlantique Nord

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published January 1996

ISBN 92-836-0021-5

Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Situational Awareness: Limitations and Enhancement in the Aviation Environment

(AGARD CP-575)

Executive Summary

This symposium was held in Brussels, Belgium 24-28th April 1995. There were 27 papers and 2 keynote addresses. Nine NATO countries contributed papers and there were 140 delegates.

The conference covered a very broad spectrum of Situation Awareness in terms of its definition, objective measurement, selection criteria, training strategy, and technology enhancements. Several incidences where fatalities had occurred as a result of loss of Situation Awareness were cited.

A significant amount of effort has been invested in defining the facets of Situation Awareness. The use of rating scales in conjunction with these defined facets identified the level of Situation Awareness achieved by individuals which could be used as selection criteria. However it was evident that whilst research into measuring the human capability for achieving and maintaining a high level of Situation Awareness has taken place there was little or no research into measuring the crew station/crew system design in terms of its ability to improve or enhance the Pilot's chance of maintaining a high level of Situation Awareness.

Training Strategy was addressed and the results highlighted that knowledge acquisition through simulation training and mission rehearsal were important contributors to acquiring and maintaining a higher level of Situation Awareness.

There were no significant advances in the cockpit technologies to improve Situation Awareness although some important research into scene linking of Head Up Display information demonstrated improved Pilot integration of the outside world with the cockpit information. Multi-modal information presentation is seen as a key element in the maintenance of Situation Awareness and the need to achieve natural intuitive interfaces was highlighted within the Open Forum Sessions.

The overall technical quality of the papers and presentations was excellent. All areas in the call for papers were targeted and the objectives of the symposium were met.

La conscience de la situation: amélioration et limitations en environnement aérien

(AGARD CP-575)

Synthèse

Ce symposium a été organisé à Bruxelles en Belgique du 24 au 28 avril 1995. En tout, 27 communications et 2 discours d'ouvertures ont été présentés par 9 pays membres de l'OTAN devant 140 participants.

La conférence a couvert un très large éventail de questions relatives à la conscience de la situation du point de vue de sa définition, de ses paramètres objectifs, des critères de sélection, des stratégies de formation et des améliorations technologiques. Suite à la perte de la conscience de la situation plusieurs cas de fatalités ont été cités.

Des efforts considérables ont été consacrés à la définition des différentes facettes de la conscience de la situation. L'emploi d'échelles d'évaluation de concert avec les facettes définies a permis d'identifier les niveaux de conscience de la situation atteints par un échantillon de sujets en vue de l'établissement de critères de sélection. Cependant, il est apparu que si des travaux de recherche ont été effectués dans le domaine de l'évaluation de la capacité humaine à atteindre et à maintenir un niveau élevé de conscience de la situation, peu ou pas de travaux de recherche ont été conduits sur l'évaluation des postes d'équipage et des systèmes/équipage du point de vue de sa capacité à aider le pilote à maintenir un niveau élevé de conscience de la situation.

La question des stratégies de formation a été abordée et il a été constaté que les résultats obtenus dans ce domaine indiquaient que l'acquisition de connaissances par le biais de l'entraînement sur simulateur et les techniques de préparation de la mission étaient des facteurs importants pour l'acquisition et le maintien d'un niveau supérieur de perception de la mission.

Aucun progrès significatif susceptible d'améliorer la conscience de la situation n'a été noté au niveau des technologies du poste de pilotage, bien que la meilleure intégration du monde extérieur avec les données cockpit par le pilote a été démontrée par certains travaux de recherche sur l'enchaînement des séquences dans le visuel tête haute. La présentation des informations multimode est considérée comme un élément clé, ainsi que la nécessité de disposer d'interfaces naturelles intuitives, ce dernier point ayant été soulevé lors des débats ouverts.

La qualité technique globale des communications et des présentations a été excellente. L'ensemble des domaines indiqués dans l'appel à communications a été couvert et les objectifs du symposium ont été atteints.

Contents

* Not available at time of printing

Preface

In order to carry out the tasks associated with meeting his mission goals within a hostile environment the pilot needs to be able to make rapid decisions and take immediate, effective and appropriate actions. To achieve this he must maintain a knowledge of his own situation and capability; and that of his adversaries. This knowledge state is termed Situation Awareness and the aim of the Symposium was to review the technological enhancements that promote good Situational Awareness and to explore the known limitations.

The Symposium focused on three main areas:

— Methods of identifying metrics for evaluating Situation Awareness in flight and in the laboratory

— Identifying instances where loss of Situation Awareness had been a clear contributory factor in aircraft losses

— Identifying areas where technologies and strategies in design, training or selection could promote good Situation Awareness.

Préface

Pour exécuter les tâches associées à la réalisation d'une mission en environnement hostile, le pilote doit pouvoir prendre rapidement des décisions et agir de façon efficace, appropriée et immédiate. Pour cela, il doit avoir en permanence connaissance de sa propre situation et des possibilités offertes, ainsi que de celles de ses adversaires. Cet état mental est appelé «la conscience de la situation». Le Symposium a eu pour objectif d'examiner les avancées technologiques permettant d'obtenir une bonne perception de la situation, ainsi que d'en déterminer les limites connues.

Le Symposium a privilégié les thèmes suivants:

— méthodologie de l'identification métrique pour l'évaluation de la perception de la situation en vol et en laboratoire

— l'identification de cas où la perte de la perception de la situation fut l'une des causes ayant contribué à la perte de l'appareil

— l'identification de domaines où le choix de technologies et de stratégies de conception, de formation et de sélection pourrait faciliter une bonne perception de la situation.

Aerospace Medical Panel

Chairman: Dr. P. VANDENBOSCH
Loriesstraat, 44
B-1500 Halle, Belgium

Deputy Chairman: LtCol A. ALNAES
Oslo Military Clinic
Oslo Mil/Akershus
N-0015 Oslo, Norway

TECHNICAL PROGRAMME COMMITTEE

Chairperson: Dr. G. Myhre
Royal Norwegian Air Force
Institute of Aviation Medicine
P.O. Box 14, Blindern
0313 Oslo, Norway

Members

Ms. J. DAVIES
British Aerospace Defence Ltd
Military Aircraft Division
Farnborough Aerospace Centre
P.O. Box 87
Farnborough, Hants GU14 6SZ
United Kingdom

Ms. S. HART
NASA Ames Research Center
MS 262-3
Moffett Field, CA 94035
USA

LCdr. D.L. DOLGIN, MC, USN
Naval Biodynamics Laboratory
13800 Old Gentilly Road, Bldg 420
New Orleans, Louisiana 70189-0407
USA

Dr. J.P. LANDOLT
D.C.I.E.M.
1133 Sheppard Avenue West
P.O. Box 2000
Downsview, Ontario M3M 3B9
Canada

Mr. J.L. FIRTH
Regional Department of Neurosurgery
and Neurology
University Hospital
Queens Medical Centre
Notthingham, NG7 2UH
United Kingdom

Dr. A. LEGER
Sextant Avionique
Rue Toussaint Catros
BP 91
33166 Saint-Médard-en-Jalles Cedex
France

LOCAL COORDINATOR

Major J.J. LECLUYSE
État-Major de la Force Aérienne
VSL/AGARD
Rue d'Evere
B-1140 Brussels, Belgium

PANEL EXECUTIVE

Major R. Poisson, CF

Mail from Europe & Canada:
Major R. Poisson, CF
AGARD/NATO
7, rue Ancelle
92200 Neuilly-sur-Seine, France

Mail from USA:
AGARD/NATO/AMP
PSC 116
APO AE 09777

Tel: (33) 1 47 38 57 60/62
Telex: 610176F
Telefax: (33) 1 47 38 57 99

TECHNICAL EVALUATION REPORT

By

Ms Jo Davies
British Aerospace (Military Aircraft) Ltd
Farnborough, Hants GU14 6YU
UK

1. INTRODUCTION

The Aerospace Medical Panel held a Symposium on '*Situational Awareness: Limitations and Enhancements in the Aviation Environment*' in Brussels, Belgium 24 - 28 April 1995. There were 27 papers and 2 keynote addresses. Two Open Forum sessions were held on the manufacturers' view on advances in Situation Awareness in current and future Civil and Military aircraft. Nine Nato countries contributed papers and there were 170 delegates.

2. THEME

This Symposium addressed an area critical to mission success and aircrew safety - Situation Awareness - which despite its common usage over the last decade lacks a fully agreed definition. The Symposium was particularly targeting research that had examined methods of identifying metrics for evaluating Situation Awareness in flight and in the laboratory, identifying instances where loss of Situation Awareness had been a clear contributory factor in aircraft losses and areas where technologies and strategies in design, training or selection could promote good Situation Awareness.

3. PURPOSE AND SCOPE

A high level of Situation Awareness forms the basis for prompt decisions and immediate, appropriate and effective actions. A low level of Situation Awareness has been identified as being the predominant cause of fatal accidents in both civil and military aviation. This Symposium aims to identify those elements both Human and Machine that contribute to achieving and maintaining a high level of Situation Awareness.

It is difficult to gauge exactly when the term Situation Awareness (SA) became common usage but it was probably in the mid eighties. Good SA became synonymous with good performance and a AGARD AMP Symposium was held in Copenhagen, Denmark in the fall 1989 (CP 478) to define conditions which may lead to loss of Pilot Situation Awareness, to evaluate methods of assessing Situational Awareness, examine methods of information presentation in the cockpit and to investigate aircrew performance and training methods.

Thus this Symposium was effectively a re-visitation of the same issues under discussion 6 years ago to facilitate an information exchange of the advances in these areas. There were 6 authors who contributed papers at both conferences (Menu and Amelberti, Taylor, Leger, Draeger, Landolt) thus providing for continuity .

4. SYMPOSIUM PROGRAMME

The Symposium programme comprised three sessions:

- Assessing Situation Awareness in Flight and in the laboratory chaired by Dolgin US and Firth UK
- Contribution of Technology to Situation Awareness chaired by Hart US and Landolt CA
- Strategies for enhancing Situation Awareness chaired by Mhyre NO and Davies UK

The first session focused on methods of evaluating Situation Awareness in flight and in the laboratory. It dealt with the research into the key psychological elements of Situation Awareness that could be used as metrics to evaluate both the ability of the man to develop and maintain situation awareness throughout their mission and the capability of cockpit designs to promote good Situational Awareness throughout the operational envelope.

It was introduced by keynote speaker Dr Grant McMillan of the US who described a large important study to develop methods of measuring Situation Awareness and methods of use of tools to select and train Pilots in maintenance of Situation Awareness.

Session 1 was followed by an Open Forum discussion - Advances in Situational Awareness in current and Future Civil Flight Deck Design. It included a video presentation of work carried out by the University of Bundeswehr into the development of knowledge-based Cockpit Assistant System for Civil Flight Decks (Onken GE) and an overview of the technology research programmes at the DCIEM (Landolt CA).

The second session concentrated on the cockpit technologies and their contribution to the maintenance and support of Situation Awareness.

The keynote speaker for this session was Dr Chris Wickens from the University of Illinois US. He gave a very comprehensive brief on all aspects of Situation Awareness including what it is, how it works and how to enhance it through display presentations.

The third session covered strategies for enhancing Situation Awareness and included advancements in measurement techniques, novel training methods and attention management.

A link session to the 1995 fall Symposium entitled the Neurological Dimensions of Situation Awareness was chaired by Firth UK and provided additional physiological and neurophysiological insights into the measurement of Situational Awareness.

The session was followed by a Panel discussion on Military Provision for Situation Awareness. John Turner UK EF2000 Project Pilot gave a presentation on the Displays and Controls philosophies on EF2000 to demonstrate how maintenance of Situation Awareness would be achieved. Captain Steele Perkins UK and Dennis Schmickley US discussed the Merlin and Apache Helicopters in a similar vein. Discussion ensued on the technological implications and the areas of risk.

5. TECHNICAL EVALUATION REPORT

There were several important recurring issues throughout the conference and this TER will draw out these issues across the papers rather than review them all individually. The issues tabled are as follows:

Can it be defined?
Can it be measured objectively?
Can it be used as selection criteria?
Can it be trained?
Can it be enhanced ?
Examples of losses of SA

Can it be defined?

Nearly all papers attempted a definition of SA. In his keynote address McMillan provided a summary table of 15 existing definitions of which the Carrol definition was adopted by the Wright Patterson Laboratory Team

' a pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats and mission and the ability to forecast, then execute tasks based on that perception.'

Crane (paper #20) argued that SA was more than simply knowledge of the environment but the successful translation into a correct employment decision.

McGuiness (paper # 7) identified a distinction between the contents of SA relating to a knowledge state and the processes that go into acquiring and using SA. He identified 5 key components Perception, Comprehension, Projection, Metacognition and Response Selection.

Wickens in his Keynote Address emphasised the need to include the response selection to performance goals as an important attribute of SA.

Firth (paper # 26) stated that neurologically it is the accurate, comprehensive, four dimensional appreciation of a situation at any one point in time.

Adam (paper # 9) was less scientific and declared that it was *' knowing what the hell's going on to know what the hell to do about it'*

Waag (paper # 8) obtained Pilot's perception of SA and the most frequently used definitions were *' building the big picture'* and *'integrating information from multiple sources'*

Can it be measured objectively ?

Having established an acceptable definition of 'what it is' the next logical question is how do you measure it.

At the previous Situation Awareness (AGARD AMP 1989) Symposium two methods of measuring SA were presented. Taylor presented the Situation Awareness Rating Technique (SART) and Endersley presented the Situation Awareness Global Assessment Technique (SAGAT).

Whilst these addressed the perception, comprehension and projection facets of SA, in order to measure the metacognition (self awareness of SA) and response selection aspects , McGuiness (paper #7) suggested that additional rating techniques are required.

Waag (paper #8) described the tools that had been developed at Armstrong Laboratories which were based upon self , supervisor and peer rating schemes. These did address the metacognition and response aspects of SA and provided an overall score based upon general traits, tactical strategies, communication and information interpretation.

Waag (paper #20) examined the relationship of performance in the simulation environment to the SA scores generated at the operational units as a validation of the measurement tool . As a positive correlation was found it was concluded that multi-ship simulation could be used as a

measurement tool. However the experience factor was dominant in the results.

It is also interesting to note from Waag's study (paper # 8) that there was a lack of correlation between self rating and Supervisor and Peer ratings. SART relies upon self rating and the aspirations of McGuiness in terms of identifying a simple rating scale 'knowing what's going on so you can figure out what to do' to measure the Pilot's subjective assessment may be problematic.

McGuiness (paper #7) described some of the limitations of post run probing and voiced caution in using recall as a measure of SA as information is only held in short term memory whilst it is relevant to the task. This aspect was concurred by Wickens who advised that there was no single measure of SA.

Taylor (paper #6) presented the work carried out at the DRA Farnborough to expand SART to include a measure of the Cognitive Compatibility construct which is identified as the sensation, perception, thinking, conceiving and reasoning elements of SA. Cognitive Compatibility is associated with goal achievement. SA manipulation is described by Hendy (paper #21) as effecting the timeliness of goal achievement and it is as a more accurate measurement of SA than the previous SART.

Mooij (paper #12) described an Eye Pointing Measuring System that could be used to provide a psychophysiological measurement of SA. It was developed as a design tool for the Columbus Workstation design activity but it was difficult to relate this to SA measurement as it was only addressing the 'perceive' process and the rest could only be inferred. In addition it was difficult to distinguish the information source in terms of whether it was on the Head Up Display or external world.

Sulc (paper #13) investigated the correlation of speech characteristics with problem situations/ emergencies. Expiration rate was found to correlate highly with low situation awareness when dealing with unexpected emergencies

The issue as to whether it was workload rather than SA that was being measured was addressed by several papers (#15, #19, #20,#21 and #22) and Hendy put across a very strong argument that the two aspects were so intrinsically linked that it was not possible to decouple the relationships He explained how time pressure could result in task shedding which then results in reduction in attention management activities. This in turn results in a reduction of information, reduction in performance as a result of poor information and an increase in workload to complete the tasks in the time available. He concurred with McGuiness that SART was actually measuring aspects of workload. The results of a simulated air traffic control experiment using both the SART metric and the TLX workload scale demonstrated very similar responses which suggests that SA and Workload are conceptually alike.

The aspect of spare capacity and its influence on the construct of SA was highlighted by several speakers. Both John Turner and Tony Steele Perkins, in the final Open Forum Session highlighted the problems of zero spare capacity and the impact that it has on decision making and attentional demand. Jensen (paper #22) described a method of measuring what was effectively spare attentional capacity using a tool called C-SAW (Continuous Subjective Assessment of Workload). The use of the Bedford Scale allows the measure of attentional demand and it is possible to use it with other uni-dimensional rating scales. The technique has been evaluated using a video tape obtained from airborne trials of a thermal imagery target

designator system integrated into a single seat cockpit. The initial results were promising.

Grau (paper #17) described work carried out at IMASAA-CERMA to model SA and identified the personalised nature of the construct of SA. He cited an example where 8 Pilots carried out the same mission in a simulator but all used different strategies to achieve the same goal. Each Pilot generally has a different knowledge state and this integrated into the comprehension of the situation accounts for the variation in responses.

Can you select for it ?

The question as to whether SA is a psychological attribute or whether it is a skill based attribute was discussed by several presenters. It was generally concurred that it was possible to select using cognitive skill capability as a criterion based on the fact that maintenance of SA required a high cognitive skill level.

Caretta (paper #3) described the work carried out by the USAF to examine methods and selection criteria for SA. Three criteria were evaluated - General Cognitive Ability, Psychomotor Ability and the personality construct of conscientiousness - as predictors of SA. The SA scores obtained from Supervisory and Peer ratings (described in paper #8) were compared with the experimental results obtained from the battery tests and demonstrated that General Cognitive Ability was the only differentiating criteria and that psychomotor skills tend to be directly proportional to flying experience.

King (paper #5) described three new tests that are administered on the Undergraduate Pilot Training Programme to test cognitive abilities and personality traits. Preliminary studies had suggested that these attributes could be predictors for aviators with superior potential for SA

Beer (paper #4) attempted to determine whether near threshold processing predicts multiple task performance. This involved performing several perceptive and cognitive tasks simultaneously. However nothing conclusive was obtained from the results. The variation in the ability of subjects to handle dual tasks may be relevant to SA selection and this aspect was included in the cogscreen battery test described in paper #5.

Ivan (paper #27) highlighted the fact that advances towards the application of colour displays in the cockpit puts more emphasis on the need to improve the colour selection tests to take account of blue yellow colour blindness in the selection criteria. He stressed that any colour vision impairment results in reduced visual range , slower reaction time, increased processing errors and thus reduced SA. The effect of lasers on colour vision was described and hence the need to continually carry out colour screening.

Freund (paper # 24) expressed concern that promising candidates (high agility measured in the psychomotoric tests) were being rejected during the selection process by virtue of EEG instability. He considered that candidates with high agility had greater potential for achieving SA due to the fact that they were faster and more intelligent with more cognitive resources and more adaptable to new situations. Experiments carried out at the German Airforce Institute of Aviation Medicine demonstrated that there was no correlation between cognitive functions measured by reaction time and changes in EEG under hyperventilation. Therefore the selection process could be relaxed in the area of EEG measurement.

Can you train for it?

The issue as to whether SA is a skill that can be acquired through training was addressed by several presenters and the overall consensus was that simulation training could accelerate the rate at which SA skills are learned and maintained.

Crane (paper #20) explored the potential of using multi-ship simulation as a training aid for teaching SA skills. The tools developed at Armstrong Laboratories were used in the measurement trials described by McMillan in his keynote address. The potential for tactical training was identified by all study participants.

Wickens, in his keynote address, emphasised the need for training in all aspects of system functionality in order to enable a mental model of system state to be maintained whilst operating out of the loop. This is of particular importance for complex systems like Flight Control Management where lack of automation awareness had been a contributory factor in several accident investigations.

Vidulich (paper #18) described the results of an interesting experiment to determine whether experience gained in attentional control using video games contributed to improved performance in a full mission simulator and whether any performance improvement was a function of improved SA. The results showed that whilst performance did improve with use of video games but there was no significant correlation with SA metrics. This is probably due to the fact that mission goals can be achieved by completely focusing on the task in hand to the exclusion of other cockpit management tasks

Maintenance of currency of SA skills was an issue raised by Spiller (paper # 23) who cited a statistic that the maximum currency window for staying at full proficiency during complex air-to- air missions was about 2 weeks.

Can you Enhance it?

The means of enhancing SA through cognitively compatible intuitive interfaces was discussed by a number of presenters.

Generally for head down display presentations there were no magic answers and Turner, in the Open Forum session challenged the designers to be more forward looking in their design solutions rather than retaining 'old hat' designs. He cited the Attitude Display Indicator and its transition from a mechanical instrument to an electronic display. (i.e. it didn't change)

Wickens discussed various approaches to maintaining SA through exocentic (outside looking in) and egocentric (inside looking out) approaches to display design but the compromise between maintaining precision vs display resolution vs wide field of view is one that cannot be resolved and there were scenarios where an egocentric approach was optimal and vice versa.

He also highlighted the fact that 3D displays which have been cited as having potential to enhance SA do in fact have deficiencies in terms of narrow field of view and positional ambiguities. Research has shown that 'highway in the sky' type displays are more intuitive for flight guidance type tasks but the field of view is too narrow to enable maintenance of SA in a global sense. They also tend to be compelling which can also result in a loss of situation awareness due to a lack of instrument monitoring .

Helmet mounted displays (HMD) was sited as a potential solution to SA maintenance by several speakers (Wickens, Adam (paper #9) and Leger (paper #14)) due to its omnidirectional

capability. In particular the cueing capability to draw the Pilot's attention to threat direction or area of concern was sited as improving awareness by Leger (paper #14).

Puleston (paper # 10) described how potential disorientation problems using helmet symbology off-boresight had been overcome with the presentation of canopy rails on the Helmet Mounted Display.

Leger (paper #14) highlighted the potential problem with fixed FLIR images projected onto the HMDs and the resultant tunnel vision syndrome. This can be overcome to a certain extent by integrating head slewable image intensifyers but update rates consistent with head movement continues to be a problem source and may contribute to disorientation effects. He described experiments focused on aiding spatial perception through the use of synthetic symbology. The discussion between conformal (scene-linked) presentations vs non-conformal (analogue and digital presentations) demonstrated that neither solution were effective in promoting good SA.

Whilst the pictorial presentations were intuitive they lacked the precision and accuracy of a digital presentation and digital presentations on their own required a higher degree of interpretation. Thus a combination of non-conformal and conformal is required.

A similar study was described by McCann (paper # 16) who carried out experiments to validate perceptual theories into the manner in which superimposed symbology is processed. In particular he was investigating runway incursions and the contributory factor of overlaid symbology. The experiments carried out demonstrated that the visual system processes visual elements with similar perceptual properties (HUD symbology equating to one group and outside world as a different group) serially with attentional resource allocated according to demand.

The experiments concluded that well defined superimposed symbology usually captured the attention and thus explained why events in the outside world could effectively go un-noticed.

Further experiments were carried out to scene-link the overlaid virtual symbology such that it is processed as the same perceptual group as the outside world. The experiments described in the paper demonstrated performance improvements with this type of approach and confirmed the findings in paper #14.

Hardiman (paper # 15) presented the results of experiments carried out at the DRA Farnborough to investigate methods of improving attitude awareness through the use of asymmetrical presentations of positive and negative pitch bars. Asymmetry was achieved through shape coding and colour coding. Significant improvements in reaction time were measured when colour and colour/shape coding was used but whether reaction time can be considered to be a measure of SA is questionable. Measurements of SA using SART for the colour coded asymmetry demonstrated that the blue/brown coding produced higher SA scores than the yellow/brown or monochrome presentation.

The results are mainly due to the compatible mental mapping of ADI displays to the blue/brown colour coding used.

Adam (paper #9) described the display technology that is currently available to promote good SA. He emphasised the need to integrate the sensor information into a form that is readily assimible. He also described the potential laser threat situation which in order to

protect the Pilot may require display projection onto an alumium surface (visor or canopy).

Turner (Open Forum) warned that the rate at which technology was advancing in terms of getting data into the cockpit was not matched with the rate of development of techniques of getting it into the Pilot's brain. Using the well worn adage there is still a danger of *'swamping him with data and starving him of information.'*

Cook (paper #19) suggested that a means of controlling the information flow to enhance SA was via different communication modes (audio and visual) such that the Pilot's cognitive load may be balanced

Diamantopoulos (paper # 25) described the physiological limitations to maintaining spatial awareness and discussed methods which could potentially alleviate the illusory disorientation experienced by aircrew. These methods included stereoscopic sound, 3D tactical displays and Head Mounted Displays.

Landolt (Open Forum) described how improvements in handling quality was being pursued at the DCIEM by means of a head mounted Usable Cueing Environment. This provided more 'heads out' time.

Both Turner (Open Forum) and Puleston (paper #10) described how the application of good design practices could be instrumental in enhancing SA. Puleston described the philosophy of work sharing in the Cobra Venom where both cockpits are identical. He also detailed the keyboard moding philosophy whereby all functions are within 1 or 2 button selections.

Turner described the information management approach on EF2000 where phase of flight moding ensured that only information pertaining to the particular phase of the mission was presented to the Pilot. He also described how DVI had transformed the cockpit in terms of increasing the time available to attend to other tasks and hence improve maintenance of SA.

Both these cases are examples of where good workload management strategies can be applied to enhance SA.

Schmickley (Open Forum) described several endeavours where technology was being brought to bear to improve workload management and increase flight safety for rotorcraft applications. He described the upgrade to the Apache Attack Helicopter and also an Obstacle Avoidance System - OASIS.

He highlighted the advances in 'associate' technology whereby crew aiding can be provided by Artificial Intelligence. He described the Rotorcraft Pilot Associate programme and in particular the Cognitive Decision Aids. The question of Crew acceptance of machine responsibility still remains an issue.

Steele Perkins (Open Forum) described the advances from the low integrity Sea King Systems to the higher integrity EH101 systems. However he questioned whether the right management information was being provided and whether there was adequate horizontal and vertical SA. He considered that further advances were required in predictive displays particularly Terrain Avoidance and Collision Avoidance Systems.

Examples of loss of SA

Loss of Situation Awareness has been a predominant cause of fatal accidents in both military and civil aviation and several presenters

cited examples where the aircraft had been lost or put in jeopardy due to Pilot error.

Cheung (paper #1) presented the results of an in depth study of all accidents over a 10 year span relating to loss of SA. The reported causes were due to unfamiliarity with environment, neglect of flight procedures, lack of effective communication among Pilots and Air Traffic Controllers, channelised attention and misjudgement.

Garcia-Alcon (paper # 2) described an incident which demonstrated the lack of communication criteria identified above. The particular incident related to a bird strike which caused injury to the 2nd Pilot in an F5 and disabled the communication system. The first pilot unaware of the 2nd Pilots injury passed flying control whilst he cleaned his visor. The aircraft descended several thousand feet before the 1st pilot finally retained full SA and took control of the aircraft at 200ft. In this circumstance the Pilot gave undivided attention to the visor cleaning task and failed to distribute his attention to the other tasks in hand.

Both Wickens and Spiller (paper # 23) identified a new criteria which is becoming increasingly more common for automated systems - that of 'complacency' or absolute trust.

Spiller (paper #23) described two incidents where trust or complacency was seen to be a contributory factor. The first was where an experienced operational crew in a Hercules C-130 were carrying out a reinforcement training sortie. The crew were flying a well worn flight plan and were motivated to achieve time on target, on speed and on height. Problems with the recovery, post drop zone, if noticed by other crew members were not communicated. The result was that the aircraft crashed into the side of a hill killing all crew members.

The second incident related to an F3 carrying out an air-air sortie off Cyprus. They were returning to base and flew into the Sea. The visual conditions were such that they were disorientated but it was concluded that the apparent lack of cross monitoring of flight instruments was due to complete trust in each others ability and judgement. In this particular case a design fault meant that the Low Altitude Warning Bug had been set to zero in order to enable the nose wheel steering warning to activate. Thus no warning of impending ground collision was given.

Wickens cited several airbus incidents where the Pilots clearly were unaware of what was going on. In the Strasbourg case an incorrect selection was made for descent rate (3000 ft/min as opposed to the desired 3 degrees flight path angle). At Kathmandu the aircraft flew into the mountain side. In both situations it is likely that the Aircrew were unaware that they were unaware. (Poor self-awareness is a function of 'mis-metacognition' defined by McGuiness (paper #20))

6. CONCLUSIONS

The overall technical quality of the papers and presentations was excellent. All areas in the call for papers were targeted and the objectives of the Symposium were met.

The only negative aspect to the Symposium was the amount of last minute changes that were made to the programme. The problem was exarcerbated by poor communication of the changes to the delegates resulting in some confusion which is somewhat ironic for a Situation Awareness Symposium.

The changes were largely due to authors withdrawing at short notice and in some cases replacement presenters were found. Generally it was seen to be preferable to contract the programme and finish on Friday at 1200 hours (as opposed to 1700 hours) rather than finish early each day.

The papers in these proceedings are in numerical order for ease of access and not necessarily in the order that they were presented.

The Open Forum Sessions provided an excellent opportunity to critically examine the extent to which the manufacturers have tackled the problem of promoting good SA in their products.

It was apparent from the manufacturers' presentations and discussions that there has been no implicit measurement of SA included in the assessment activity of crew station design. Simulation plays an important role in providing early visualisation of the design goals and generally the 'fit for purpose' criteria is derived from subjective aircrew assessment.

Where affordability is a design driver, managing cost effectiveness of the product becomes a key activity. The ability to carry out a cost benefit analysis of the contribution of the MMI to the overall system design is still lacking. In particular the ability to justify changes on the grounds of improvement to SA is still very difficult due to the lack of techniques and tools.

The symposium demonstrated the significant advances in both understanding SA and measuring the Human elements that contribute to achieving, maintaining and repairing SA. However there is still further work to achieve similar advances for the Machine elements in order to differentiate cockpit integration that promotes good SA from those that don't.

Whilst the technology does exist to present multi-modal information to the Crew there is still a danger that the right information is not immediately accessible to enable the Pilot to be responsive and reactive to his situation in an accurate and rapid manner. There was no real evidence that a true synergy between the Man and Machine has been achieved.

One very clear message is that accidents of the type described in the Symposium will continue to occur unless improvements to Crew / System Integration can be effected to guarantee an acceptable level of Situation Awareness throughout the operational envelope.

7. RECOMMENDATIONS

The Symposium addressed the current state of the art in terms of limitations and enhancements of Situation Awareness in the aviation environment. Several areas were identified where it is considered that further research is required by the NATO communities in order to effect improvements in Human System Integration:

1. Research into the measurement aspects of SA needs to continue in order to provide the system designers with objective assessment criteria which they are currently lacking

2. Information presentation still appears to be a problem and in order to provide natural intuitive interfaces it is essential that the endeavours in the area of cognitive compatibility be pursued.

3. Knowledge acquisition in both tactical strategies and on-board system understanding forms an important element of SA and research should continue into methods of enhancing knowledge acquisition by such means as mission rehearsal aids and system trainers.

4. There are still too many examples of poor integration of automation in today's cockpits which adversely affect SA and further research is required into methods of achieving a more synergistic implementation.

5. The relationships between Workload, SA and Performance still requires further exploration to fully understand their mutual dependencies.

EVALUATING PILOT SITUATIONAL AWARENESS IN AN OPERATIONAL ENVIRONMENT

Dr. Grant R. McMillan
Armstrong Laboratory
Human Engineering Division
AL/CFHP
2255 H Street
Wright-Patterson Air Force Base OH 45433-7022
USA

Lt Col James Bushman
Armstrong Laboratory
Technical Training Research Division
AL/HRT
7909 Lindbergh Drive
Brooks Air Force Base TX 78235-5352
USA

Maj Carol Lynn A. Judge
AF Studies and Analysis Agency
Directorate of Modeling, Simulation and Analysis
AFSAA/SAGT, Room 1D380
1570 Air Force Pentagon
Washington DC 20330-1570
USA

1. SUMMARY

The Chief of Staff of the U.S. Air Force requested a major study with the following objectives: (1) Develop measures of pilot situational awareness (SA); (2) Identify tools for selecting pilots most likely to develop good SA; (3) Identify tools for training SA. This request mobilized a massive study of over 200 F-15 pilots which was conducted largely in the field. The study included the development and validation of a set of SA ratings scales designed for use in an operational environment. A transportable, computer-based SA Aptitude Battery was also developed and used at F-15 squadrons in the continental United States, Alaska, and Japan. Finally, a subgroup of over 40 pilots was evaluated in a high-fidelity simulation environment which permitted detailed testing of numerous behavioral aspects of SA. This paper describes the methodology, findings, and lessons learned from this study.

2. BACKGROUND

The United States Air Force recently completed a major study of fighter pilot situational awareness (SA) in response to a request from the Chief of Staff. Figure 1 shows that request from General McPeak. When the Armstrong Laboratory's Situational Awareness Integration (SAINT) team received the memo, we determined that he was asking questions in three areas: questions with respect to the measurement of SA, questions with respect to the learning of SA skills, and questions with respect to the possibility of selecting pilots, early in the flying training process, who are likely to demonstrate good SA in the air combat environment. Since almost all of the previous SA research had involved highly controlled experiments in laboratory environments, General McPeak's questions represented a real challenge. This paper describes the study that was conducted to address this challenge, summarizes the study findings, and discusses lessons learned in the operational testing process. The results of the study are presented in detail by Carretta and Ree (Ref 1), Waag and Houck (Ref 2), and by Waag, Houck, Greschke and Raspotnik (Ref 3), all of which are included in this volume.

3. DEFINITIONS OF SITUATIONAL AWARENESS

The situational awareness construct continues to have a major impact on the aviation research community, despite the fact that there is no agreed upon definition. Dominguez (Ref 4) compares 15 published definitions of SA (Table 1) and discusses several important dimensions on which they differ. One of these dimensions is the extent to which SA is viewed as a skill or ability akin to "the right stuff", or as a state of mind which may or may not be achieved in a particular flight situation. Both concepts are important for a comprehensive understanding of SA since significant mental ability is necessary to support the assessment and interpretation of situations, and situational factors such as stress and fatigue can undermine SA no matter how impressive one's mental abilities.

General McPeak's memorandum poses questions pertinent to the ability issue such as "Can it be learned?" and "Is it a sex related capacity?". The ability notion is reflected in the following definition which was developed by an Air Staff SA working group (Ref 5):

A pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception.

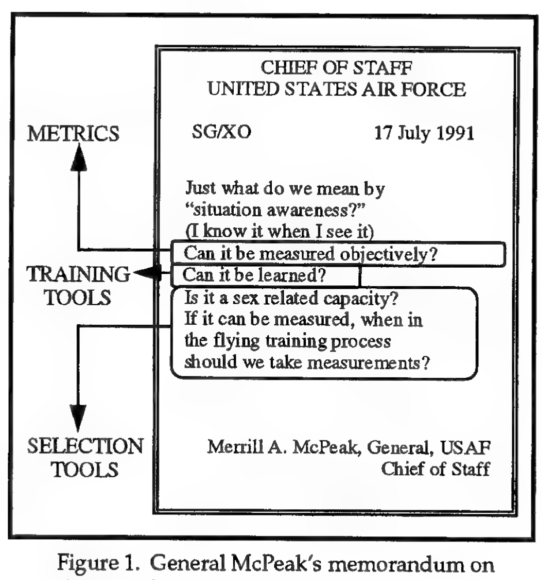


Figure 1. General McPeak's memorandum on situational awareness.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

Table 1. Definitions of Situational Awareness (From Dominguez, 1994)

Definitions	Source
Conscious awareness of actions within two mutually embedded four-dimensional envelopes.	Beringer and Hancock (Ref 6, p.646)
A pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission and the ability to forecast, then execute tasks based on that perception.	Carroll (Ref 5)
The ability to extract, integrate, assess, and act upon task-relevant information is a skilled behavior known as 'situational awareness.'	Companion, Corso, Kass, & Herschler (Ref 7)
The accurate perception of the factors and conditions that affect an aircraft and its flight crew.	Edens (Ref 8, p. 7); Schwartz (Ref 9)uses this definition with "during a defined period of time" at the end.
The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.	Endsley (Ref 10, p.1-3)
The knowledge that results when attention is allocated to a zone of interest at a level of abstraction.	Fracker (Ref 11, p. 102)
The pilot's overall appreciation of his current 'world.'	Gibson and Garrett (Ref 12, p.7-1)
One's ability to remain aware of everything that is happening at the same time and to integrate that sense of awareness into what one is doing at the moment.	Haines and Flateau (Ref 13, p. 43)
<u>Where</u> refers to spatial awareness. . . <u>what</u> characterizes identity awareness, or the pilot's knowledge of the presence of threats and their objectives, [as well as] engine status and flight performance parameters. <u>Who</u> is associated with responsibility, or automation awareness; that is knowledge of 'who's in charge.' Finally, <u>when</u> signifies temporal awareness and addresses knowledge of events as the mission evolves.	Harwood, Barnett, and Wickens (Ref 14, p. 316)
The ability to envision the current and near-term disposition of both friendly and enemy forces."	Masters, McTaggart, and Green (Ref 15, p.5); Stiffler (Ref 16)
Awareness of conditions and threats in the immediate surroundings.	Morishige and Retelle (Ref 17, p. 92)
The ability to maintain an accurate perception of the surrounding environment, both internal and external to the aircraft as well as to identify problems and/or potential problems, recognize a need for action, note deviations in the mission, and maintain awareness of tasks performed.	Prince and Salas; cited in Shrestha et al., (Ref 18, p.10)
[Situational awareness] means that the pilot has an integrated understanding of factors that will contribute to the safe flying of the aircraft under normal or non-normal conditions.	Regal, Rogers, and Boucek (Ref 19, p.65)
Situation awareness refers to the ability to rapidly bring to consciousness those characteristics that *evolve* during flight.	Wickens (Ref 20, p.2)
The pilot's knowledge about his surroundings in light of his mission's goals.	Whittaker and Klein (Ref 21, p.321)

Interviews with a number of Air Force fighter pilots suggest that they share the view that ability is important, and that they also have a great deal of difficulty separating SA from pilot expertise. For them, SA and fighter pilot ability go hand in hand.

The Armstrong Laboratory team adopted the above definition, although we were aware that there is little consensus in the SA community. The approach used for measurement of SA in the operational environment reflected this skill-ability perspective. The SA rating scales required pilots to rate their subordinates or peers on a variety of behavioral factors which they had observed over *a significant period of time*. This assumed that SA would be a consistent individual characteristic demonstrated in most tactical situations. Although one

might debate the wisdom of this approach, testing a large number of pilots in an operational environment afforded few other options. The portion of this study in which we evaluated SA in a controlled simulation environment, on the other hand, permitted situation-specific measurements. In fact, a key question was the extent to which the operational SA measures would correlate with the SA measurements taken in the simulator.

Another controversial feature of this definition is its inclusion of *task execution* as an element. Only one other definition in Table 1 includes this behavioral factor. Although it is unusual, it does reflect the view of the operational forces. (The Air Force working group which developed the definition was primarily composed of pilots.) It can be argued that including task execution in an SA definition further blurs the distinction between situational awareness and pilot expertise. The merits of carefully separating or tightly linking perception and action will continue to be debated in this and other theoretical contexts. We choose to accept this tight linkage because it reflects the views of the pilots who requested and participated in this significant SA study.

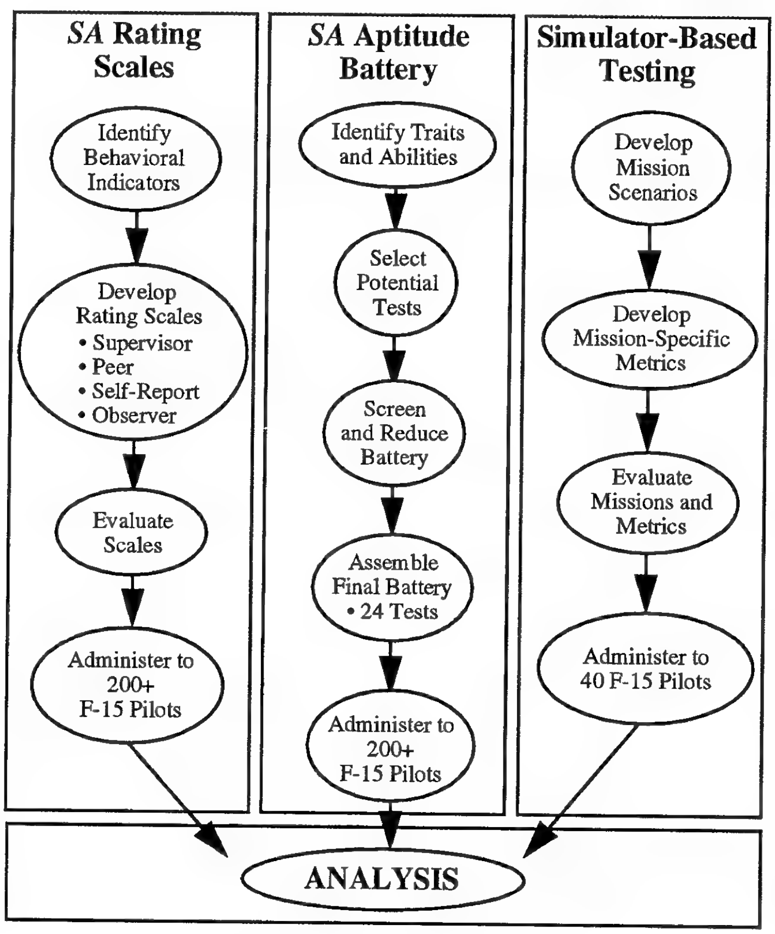


Figure 2. Structure of the Armstrong Laboratory situational awareness study.

4. STUDY METHODOLOGY

Figure 2 illustrates the overall structure of the study. To measure SA in an operational environment four rating scales were developed: a supervisory rating form, a peer rating form, a self-report scale, and an observer form designed to evaluate SA in simulated air combat. The observer form was used in the simulator-based testing shown on the right side of Figure 2.

These SA rating scales were designed to make the "I know it when I see it" process (a phrase often used to describe SA) systematic, repeatable, and reliable. The 31-element scales sampled behaviors such as management of flight communication, selection of targets, selection of weapons, system management, and interpretation and integration of information. The elements were based on an F-15 air combat mission task analysis conducted by Houck, Whittaker and Kendall (Ref 22). This analysis was designed to identify the critical activities required for successful completion of an air combat mission. The behavioral elements identified in this analysis were reviewed by a subject matter expert to select those which are essential to SA. Additional criteria were that the behavioral elements must be observable in day-to-day squadron operations and subject to evaluation by other pilots. Concise definitions of each item were developed with the assistance of an experienced fighter pilot. The scales were pre-tested at Tyndall Air Force Base to ensure that pilots could use them easily and that they understood what they were being asked to rate. Trained teams supervised the administration of the scales to over 200 pilots in F-15 units at Eglin, Langley, Elmendorf, and Kadena Air Force Bases. Each pilot had ratings from over 30 people, including the peer and supervisory evaluations.

The second element of the program was the SA Aptitude Battery. It consisted of a set of computer-based tests designed to measure skills and abilities that were believed to be important for the development and maintenance of SA. Multiple empirical studies (see Ref 1 for a review) have identified three general factors that are valid for predicting job performance in many contexts. These factors are general cognitive ability, psychomotor skill, and the personality construct "conscientiousness". Based upon these findings, and on previous SA research, tests were selected to measure these three factors. The specific tests measured elements such as working memory, spatial processing, reasoning,

extroversion, agreeableness, conscientiousness, multi-limb coordination, control precision, reaction time, etc. Most of the tests were previously developed, but several had to be refined for this application. The resulting battery was thoroughly pre-tested with basic trainees and pilots. Based upon this pre-testing, the battery was reduced to 24 tests that could be administered in two, two-and-one-half-hour sessions. The personal computer-based battery was housed in portable testing booths that were transported to each fighter wing or squadron. The battery was administered by trained test teams to the majority of the pilots for whom we had SA rating scale data.

The third element of the program addressed SA measurement in a controlled simulator environment. It employed Armstrong Laboratory simulator facilities located principally in Mesa, Arizona, USA. Forty pilots were chosen from the above set of over 200 to serve as experimental subjects. The pilots were selected to represent a broad range of operational SA, as indicated by their squadron ratings. In addition to the experimental pilots, the simulation employed live wingmen, ground control intercept (GCI) operators, airborne warning and control system (AWACS) operators (via a link to a simulation at Brooks Air Force Base, Texas, USA), two manned threat aircraft, and up to four automated threats and four attack aircraft. The highly-scripted scenarios ranged from 1 versus 2 up to 2 versus 6. The experimental pilots were tested one at a time and flew 36 engagements over 5 days. Scenario difficulty was varied by manipulating the amount of surface-to-air missile activity, communications jamming, GCI and AWACS support, and the number and type of threats.

A variety of dependent variables were collected in the simulator, including objective measures related to kills, time to acquire targets, and eye movement data. The subjective evaluations, using the observer form of the SA rating scale, proved to be a powerful measurement technique. The two observers were highly trained ex-Air Force pilots who were blind to the experimental pilots' squadron ratings. The rating process involved independent observation and rating of each sortie by each observer. This was followed by subject debriefings and the development of consensus ratings after each simulator test.

5. SUMMARY OF STUDY FINDINGS

The data collection process was completed in 14 months. Good data sets were available for the following numbers of pilots:

Squadron SA Ratings	-	238 pilots
SA Aptitude Battery	-	171 pilots
Simulator SA Scores	-	40 pilots

(Some pilots were unable to complete the SA Aptitude Battery because of scheduling conflicts and some data were lost or contaminated due to equipment malfunctions.)

The squadron SA ratings were analyzed to develop a criterion measure of operational SA. This process is described in detail by Waag and Houck (Ref 2), and resulted in a criterion which accounted for 92.5% of the variance in the supervisory and peer ratings. In addition, the peer and supervisory ratings were highly correlated ($r = .87$) indicating high agreement among the different raters. The self-report ratings showed much lower correlations with the peer and supervisory ratings ($r = .58$ and .50, respectively) and were not included in the criterion measure. The resulting distribution of these criterion scores is shown in Figure 3. The utility of this SA measure is further indicated by the fact that there were no important differences in the average scores of the eleven F-15 squadrons that participated in the study, suggesting that they all used the rating scales in a similar fashion. Observer ratings in the simulator also proved to be a highly effective technique for measuring SA in this more controlled environment (Ref 3). Correlation of the squadron and simulator SA ratings (Figure 4) suggests that performance in these two environments is related.

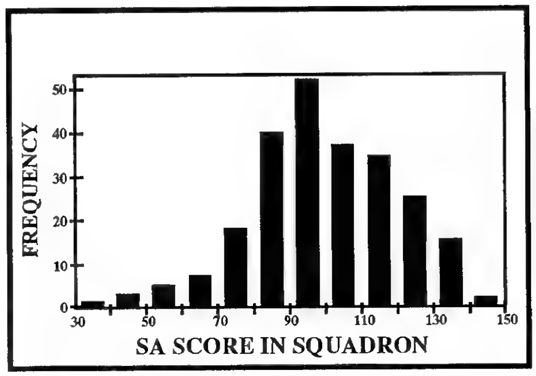


Figure 3. Distribution of operational SA scores based on squadron ratings.

The second question raised by General McPeak was "Can SA be learned? This question was not directly addressed by the current effort, since it was not a longitudinal training study. Nevertheless, several lines of converging evidence suggest that SA can be enhanced by training. First, analysis of the squadron ratings showed that pilots who had training such as Fighter Weapons School and Red Flag, Green Flag, or Maple Flag exercises had significantly higher SA scores. Second, analysis of pilot performance during the week-long simulator testing showed evidence of improvements in SA. Finally, pilot

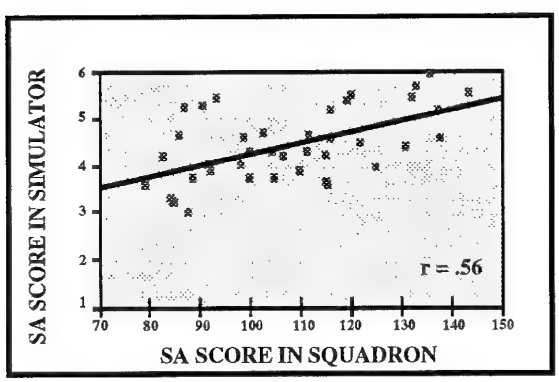


Figure 4. Correlation of squadron and simulator SA scores.

opinion data collected at the completion of the simulator testing strongly supported the SA training benefit of interactive, multi-ship simulation (Ref 3).

The third question addressed by the study was the extent to which the operational SA ratings could be predicted from the aptitude battery scores. In conducting this analysis (see Ref 1 for details), we observed that the best predictor of squadron SA rating was the number of F-15 flying hours. Consequently, F-15 hours was used as a control variable in subsequent regression models which evaluated various SA aptitude battery scores. These regression models demonstrated that cognitive ability was predictive of the criterion, as expected from other job performance studies. On the other hand, psychomotor scores and the personality trait "conscientiousness" were not predictive of operational SA.

6. LESSONS LEARNED

The SA study summarized above represents the largest effort of its kind, to date. The key factor in accomplishing this study was the commitment of top Air Force leaders. This began with the direction of General McPeak and included the leadership of the Air Force Materiel Command and the Human Systems Center, our parent organizations. In addition, the commitment of the Air Combat Command, including the wing and squadron commanders, was essential to the success of this program. Operational pilots view any such testing with suspicion and as a burdensome addition to their busy schedules. Nevertheless, operational testing is critical to the development of useful SA measurement, selection, and training tools, and it can only be accomplished if operational leaders endorse the program.

Elements that were important to the timely completion of data collection were the use of on-site testing methods, well-trained test administration teams, and flexible pilot scheduling. A strong and committed squadron point-of-contact (POC) was also important. Data collection required only about one-half the time for squadrons with an effective POC. A key concern of the pilots and squadron leadership was the confidentiality of these potentially sensitive data. Subject coding and careful separation of data files and subject code files were employed to address these concerns.

Pilot comments on the testing methods provided highly useful feedback. The SA rating scales were well-received, including the supervisory and peer rating process. Pilots reported little difficulty in using any of the rating instruments. The simulator testing was also well-received, including the tactical scenarios and the observer rating process. User opinion data were gathered on the perceived training benefit of this multi-ship interactive simulation. The study participants consistently rated the simulation as beneficial for all levels of pilot experience. The SA aptitude battery, on the other hand, was not as positively perceived. Most pilots felt that five hours of testing was too long and had difficulty identifying with the context-free nature of this battery. Despite these concerns, good data were collected on a majority of the participants. Based upon the results of this study, future SA test batteries can be significantly shorter in length.

The advantages of using subjective techniques to quantify complex constructs, such as SA, have been discussed by Hennessy (Ref 23). His report was one factor that motivated us to employ structured ratings and trained observers in this study. The squadron rating process proved to be a highly effective technique for SA measurement. With approximately 30 independent ratings for each subject, perceptual errors and biases were well controlled. The use of trained observers in the simulator testing offered numerous advantages over purely objective approaches, as well. These advantages included rapid data analysis, the ability to disregard simulator deficiencies when evaluating pilot performance, the ability to identify "simulator gaming" behavior which tends to distort objective measures, and the ability to provide insightful interpretations of observed behavioral events. In addition, we are finding the observer ratings and comments to be an excellent guide in our search for sensitive and reliable objective measures. Clearly, SA evaluations that involve complex tactical scenarios should include structured expert observations as part of the measurement process. Objective measures alone are not yet up to the task.

7. REFERENCES

1. Carretta, T.R., and Ree, M. J. (1995). Determinants of situational awareness in U.S. Air Force F-15 Pilots. In AGARD-CP-575, *Situation Awareness: Limitations and Enhancements in the Aviation Environment.* Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development.

2. Waag, W.L., and Houck, M.R. (1995). Development of criterion measures of situation awareness for use in operational fighter squadrons. In AGARD-CP-575, *Situation Awareness: Limitations and Enhancements in the Aviation Environment.* Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development.

3. Waag, W.L., Houck, M.R., Greschke, D.A., and Raspotnik, W.B. (1995). Use of multiship simulation as a tool for measuring and training situation awareness. In AGARD-CP-575, *Situation Awareness: Limitations and Enhancements in the Aviation Environment.* Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development.

4. Dominguez, C. (1994, June). Can SA be Defined? In M. Vidulich, C. Dominguez, E. Vogel, & G. McMillan (Eds.), *Situation Awareness: Papers and Annotated Bibliography* (Tech. Rep. No. AL/CF-TR-1994-0085, pp. 5-15). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

5. Carroll, L.A. (1992, March). Desperately seeking SA. *TAC Attack* (TAC SP 127-1), *32*(3), 5-6.

6. Beringer, D.B., and Hancock, P.A. (1989). Exploring situational awareness: A review and the effects of stress on rectilinear normalization. In *Proceedings of the Fifth International Symposium on Aviation Psychology* (Volume 2, pp. 646-651). Columbus. OH: The Ohio State University.

7. Companion, M.A., Corso, G.M., Kass, S.J., and Herschler, D.A. (1990, January). *Situational awareness: An analysis and preliminary model of the cognitive process* (IST-TR-89-5). Orlando, FL: University of Central Florida, Institute for Simulation and Training.

8. Edens, E.S. (1991, April). *Individual differences underlying pilot cockpit error.* (Tech. Report No. DOT/FAA/RD-91-13). Washington, DC: US Department of Transportation. (AD-A236107)

9. Schwartz, D. (1993, May). Training for situational awareness. *Air Line Pilot, 62*(4), 20-23.

10. Endsley, M.R. (1990, April). A methodology for the objective measurement of pilot situation awareness. In AGARD-CP-478, *Situational Awareness in Aerospace Operations* (pp. 1-1 to 1-9). Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development. (AD-A223939)

11. Fracker, M.L. (1990, April). Attention gradients in situation awareness. In AGARD-CP-478, *Situational Awareness in Aerospace Operations* (pp. 6-1 to 6-10). Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development. (AD-A223939)

12. Gibson, C.P., and Garrett, A.J. (1990, April). Towards a future cockpit - The prototyping and pilot integration of the mission management aid (MMA). In AGARD-CP-478, *Situational Awareness in Aerospace Operations* (pp. 7-1 to 7-9). Neuilly Sur Seine, FR: Advisory Group for Aerospace Research & Development. (AD-A223939)

13. Haines, R.F., and Flateau, C. (1992). *Night flying.* Blue Ridge Summit, PA: TAB Books.

14. Harwood, K., Barnett, B., and Wickens, C. (1988, April). Situational awareness: A conceptual and methodological framework. In *Proceedings of the Psychology in the Department of Defense Eleventh Symposium* (Tech. Report No. USAFA-TR-88-1, pp. 316-320). Colorado Springs, CO: US Air Force Academy. (AD-A198723)

15. Masters, R.M., McTaggart, T.E., and Green, G.L. (1986). *Air-to-air F-15/F-16 identification analysis methodology - Management summary.* (Tech. Report No. ASD-TR-87-5004) Wright-Patterson Air Force Base, OH: Combat Identification System Program Office. (AD-B108209)

16. Stiffler, D.R. (1987, April) *Exploiting situational awareness beyond visual range.* (Report No. 87-2370) Maxwell Air Force Base, AL: Air Command and Staff College.

17. Morishige, R.I., and Retelle, J. (1985, October). Air combat and artificial intelligence. *Air Force Magazine, 68*(10), 91-93.

18. Shrestha, L.B., Prince, C., Baker, D., and Salas, E. (In press). Understanding situation awareness: Concepts, methods, and training. To appear in W.B. Rouse (Ed.), *Human/technology interactions in complex systems.*

19. Regal, D.M., Rogers, W.H., and Boucek, G.P. (1988). *Situational awareness in the commercial flight deck: Definition, measurement, and enhancement* (SAE Technical Paper Series No. 881508). Warrendale, PA: Society of Automotive Engineers.

20. Wickens, C.D. (1992, December). Workload and situation awareness: An analogy of history and implications. *INSIGHT: The Visual Performance Technical Group Newsletter, 14*(4), 1-3.

21. Whittaker, L.A., and Klein, G.A. (1988, April). Situation awareness in the virtual world: Situation assessment report. In *Proceedings of the Psychology in the Department of Defense Eleventh Symposium* (Tech. Report No. USAFA-TR-88-1, pp. 321-325). Colorado Springs, CO: US Air Force Academy.

22. Houck, M.R., Whittaker, L.A., and Kendall, R.R. (1993). *An information processing classification of beyond-visual-range intercepts* (Tech. Report No. AL/HR-TR-1993-0061). Brooks Air Force Base, TX: Armstrong Laboratory.

23. Hennessy, R.T. (1990). Practical human performance testing and evaluation. In H.R. Booher (Ed.), *Manprint.* New York, NY: Van Nostrand Reinhold.

LOSS OF AVIATION SITUATION AWARENESS IN THE CANADIAN FORCES

Bob Cheung Ph.D., Ken Money Ph.D., Paul Sarkar B.Sc.

Defence and Civil Institute of Environmental Medicine
P.O. Box 2000, North York, Ontario
Canada
M3M 3B9

SUMMARY

The ability of a pilot to maintain situation awareness has been recognised as crucial to mission success for some time. Situation awareness has been used to refer narrowly to a combination of tactical awareness and spatial orientation. However, situation awareness is the accurate perception and comprehension of a number of factors and conditions that could affect the aircraft and the air crew within a period of time. The present study concentrates on accidents and incidents in which there was a loss of situation awareness excluding spatial disorientation.

Each accident and incident were systematically reviewed to assess the role of situation awareness. Loss of situation awareness has been implicated in many close calls and accidents. A total of 64 mishaps between 1982-1993 were found to be related to loss of situation awareness in the Canadian Forces (CF) and it appeared throughout all mission and aircraft types.

A focused and structured training program in managing cockpit resources and in maintaining attention would assist air crew in identifying conditions where situation awareness could potentially be lost and where appropriate strategies could be used to avoid the loss or to deal with the loss. Such training could be implemented through real-time man-in-the-loop flight simulator training of pilots in various flight scenarios. Similar training could also improve the performance and efficiency of air traffic controllers.

LIST OF ABBREVIATIONS

ACM	air combat manoeuvre
AGL	above ground level
ATC	air traffic control
CF	Canadian forces
G	gravity
IFR	instrument flight rules
kias	knots indicated air-speed
LH	left hand
LSA	loss of situation awareness
MDA	minimum descent altitude
RH	right hand
RPM	revolutions per minute
SA	situation awareness
SD	spatial disorientation
SO	spatial orientation
TSA	tactical situation awareness
VFR	visual flight rules

INTRODUCTION

Situation awareness (SA) is considered to be a crucial prerequisite for the safe operation of complex dynamic systems especially in aviation. Currently it is a fashionable concept among students of cockpit automation and pilot performance. The definition of situation awareness varies considerably. There is as yet no satisfactory definition of SA or boundary to constrain the concept. Historically, SA has referred to tactical situation awareness, i.e. how pilots gain awareness of the enemy before they gain awareness of themselves and how pilots devise methods to complete the mission. In relation to human cognition, SA was defined as the perception of the relevant elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (1). The elements of this definition were explicitly defined for air-to-air tactical missions. Fracker (2) defines SA as the knowledge that results when attention is allocated to a zone of interest at a given level of abstraction. In human/machine systems, SA is defined as the conscious awareness of actions within two mutually

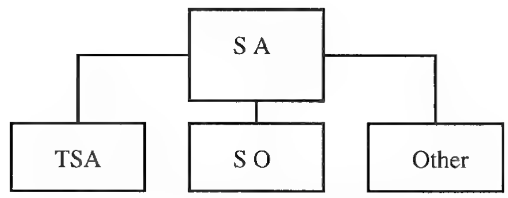
Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

embedded envelopes with the inner envelope consisting of unaided sensory space of the operator and the outer envelope consisting of information available to the operator via remote sensing (3). Despite the importance and popularity of SA, little progress has been made with respect to better understanding and successfully manipulating the phenomenon from an operational point of view.

There have been numerous task analysis studies investigating the frequency and severity of spatial disorientation (SD) implicated accidents and incidents (4, 5, 6, 7, 8). However, data on the prevalence of loss of situation awareness (LSA) are very scanty. There were very few studies which directly dealt with the incidents and accidents due to the loss of situation awareness and some of the scientific literature tends to deal with situation awareness and spatial disorientation synonymously (9, 10). Part of the reason for such a discrepancy is that some investigators have included LSA in the list of illusions one could encounter when spatially disorientated (11). Others attempted to overcome the classification problems by grouping accidents in a combined class called SD/LSA (12). There are very few operational definitions of SA (6, 13) and a commonly accepted operational definition of SA is lacking.

Operationally, the term situation awareness in the aviation environment is used more appropriately to refer to the awareness of the large group of factors that are important in keeping the aircraft safe from hazardous situations or potentially dangerous flight paths. These factors include geographical location, attitude, weather, tactical environment, weapons capabilities, individual capacities, effective communication, administrative constraints, adherence to proper flight rules, and also spatial orientation (SO). This hierarchical structure of spatial orientation as a part of SA has been proposed by Gillingham (6) and more recently by Navathe and Singh (14). In an attempt to derive an operational definition of spatial disorientation, Navathe and Singh limited loss of SA as a psychological limitation/overload, a condition wherein the aircraft enters a dangerous flight path as a result of central error due to illusion, error of judgment, lack of information or preoccupation (14). SA is intended to be the broader term encompassing more than spatial disorientation (SD) references and including the cognitive process as its principal dimension. In other words, SD is one kind of loss of SA, but loss of SA could also be due to factors other than SD. (See Fig. 1).

Figure 1: Hierarchical Structure of Situation Awareness



The objective of this study is to describe incidents and accidents related to loss of SA due to factors other than spatial disorientation according to the aforementioned operational definition. A separate study on spatial disorientation implicated accidents in the Canadian Forces have recently been presented elsewhere (4). An attempt will be made to classify LSA according to the various possible cause factors. Information was gathered concerning the genesis of loss of situation awareness so that training and research efforts could be appropriately applied.

METHODS

Narratives regarding accidents and incidents between 1982 and 1993 were obtained from ACAIRS (Aircraft Accident Incident Reporting System) of the CF Directorate of Flight Safety. Each accident and incident was systematically reviewed to assess the role of loss of situation awareness. Within the Canadian Forces, formal inquiries were held only for Category A accidents. In addition to investigating the reported cause factors, specific information was collected for each accident and where possible for each incident. This specific information included: pilot experience, aircraft type, mission profile, time of day, weather and terrain of the flight path. This information was tabulated and summarised for analysis.

The CF Handbook of Flight Safety defines an accident as an event in which an aircraft is missing or in which there is A, B or C category damage, or a person receives fatal or serious injury. A Category A accident is when an aircraft is destroyed, declared missing, or damaged beyond economical repair. A Category B accident is when the aircraft must be shipped to a contractor or depot-level facility for repair. A

Category C accident is when the aircraft must be flown to a contractor or depot-level facilities for repairs; repairs are carried out by a mobile repair party; or a major component has to be replaced. An incident is defined as an event in which there is D category damage, when damage to any component of the aircraft can be repaired within field-level resources or a person receives minor injury. A Category E incident is when there is no aircraft damage, but accident potential exists. (15). These definitions were adopted for this study.

RESULTS

Between 1982 and 1993, a total of 64 mishaps were found to be related to loss of situation awareness excluding spatial disorientation, 11 in helicopters and 53 in fixed wing aircraft. There were 3 Category A, 2 Category B and 1 Category C accidents. The rest were classified as incidents including 5 in Category D and 53 in Category E. The time of day when and the terrain of the flight path where the accidents or incidents occurred appeared not to play a role in the mishaps.

Aircraft type

Accidents and incidents related to loss of situation awareness span across all aircraft types as illustrated in Table 1.

Table 1: Situation Awareness Accidents and Incidents by Aircraft Type

Aircraft Type	SA Accidents and Incidents
CF 18 (Hornet)	18
CF 5 (Freedom Fighter)	12
CT 114 (Tutor)	12
CT 133 (Silver Star)	6
CH 124 (Sea King)	3
CH 135 (Twin Huey)	3
CP 140 (Aurora)	3
CH 136 (Kiowa)	2
CC 115 (Buffalo)	1
CH 113 (Labrador)	1
CH 139 (Jet Ranger)	1
CH 147 (Chinook)	1
CP 121 (Tracker)	1

Mission profiles

About 30% of the mishaps occurred within the air space centred around the runway that extends from the ground to 1,000 feet (300 m.) above ground level (AGL). There were 14 SA-related mishaps during landings, 5 during takeoffs, 8 during formation flying, 18 during air combat manoeuvres, 17 during routine flight training, and 2 during aerobatics manoeuvres.

Weather

The occurrence of LSA was not limited to extreme weather condition. A number of near misses and a fatal collision with objects occurred on very good VFR days. There were three incidents involving the degradation of weather beyond forecasted levels. A solo student pilot who did not have valid instrument ratings was airborne at the time of an amended forecast on a clearhood mission that could have necessitated diversion. Airborne pilots who were aware of the worsening weather situation did not pass the information to allow for a timely forecast amendment. Another case involved a pilot who chose to continue and descend below MDA (Minimum Descent Altitude) to gain visual reference with the airport despite the forecast weather was below MDA for the planned approach. The pilot landed the aircraft only after 3 attempts. In another case, no warning was given about severe wind shear and turbulence, and visibility was extremely limited due to heavy rain and darkness at night; the pilot encountered moderate mixed icing and severe wind shear through the descent and final phase of the approach.

Flying experience

In general, the flying experience of the pilots played a minor role in most cases. Pilots of all ages and of varying amounts of flying experience are susceptible to loss of situation awareness. In those incidents where experience played a role, the inexperienced younger pilots were not familiar with all the possibilities and limitations of the specific aircraft.

DISCUSSION

Circumstances and Causal Factors

For any accident investigation, it is difficult to attribute one single cause factor for the mishap, especially in modern fighter aircraft with advanced technology and difficult mission requirements. When there is a loss of life, it is often especially difficult to be sure of the cause. Accidents are usually the result of a chain of events that culminate in the mishap. There is seldom one overpowering cause, but rather a

number of contributory factors or errors. The cause factors assigned should by no means be treated as the only cause, but as contributory factors under the circumstances. All of the accidents and incidents that we reviewed involved one or more of the following as tabulated in Table 2:

Table 2: Contributory Factors and Related Problems

Contributory Factor	Problem Experienced
Geographical Location	(i) Unfamiliar with environment away from home base. (ii) Lack of awareness of altitude.
Weather	Unforeseen weather condition.
Individual Capacity	(i) Limitation of personal capacity. (ii) Inattention. (iii) Distraction. (iv) Channelised attention.
Adherence to Proper Flight Rules	(i) Failure to maintain adequate clearance around aircraft. (ii) Failure to maintain instrument scan. (iii) Failure to observe instructions from tower.
Administrative Constraints	With available flying hours steadily declined, the inexperienced younger pilots would not be as familiar with all the possibilities and limitations of the specific aircraft as desired.
Effective Communication	(i) Use of non-standard procedures or instruction. (ii) Lack of effective communication among pilots and aircrew. (iii) Lack of effective communication between pilots and air traffic controllers.

Consequences of loss of SA

Loss of situation awareness resulted in mid-air collision or collision with a ground-based object, near misses, high G overstress, ground/water impact, undercarriage overspeed, departure from controlled flight and unawareness of low fuel state.

Collision

Collision or contact with ground-based objects commonly occurred when air crews failed to maintain adequate clearance around the aircraft, i.e. rotor blades or undercarriage striking tree tops. In one case, an unoccupied seat pack was improperly secured and departed the aircraft through the canopy glass when the pilot executed a right bank. The seat pack broke the glass.

Near miss

A number of near misses occurred during multi-bogey ACM (Air Combat Manoeuvre) training during landing after the formation break-up. Other near misses occurred as a result of unsafe overshoot procedures during landing. The lack of awareness of the layout of an away base where landing was to be made resulted in landing on taxi way. Quite a number of near misses occurred in very good VFR conditions during the approach to landing or the landing phase itself.

Overstress

Most of the G overstress cases occurred when pilots unknowingly failed to maintain attitude, allowing the nose to drop too far and causing overstress of the aircraft during subsequent aggressive pull-up. Other nose low situations occurred during various aircraft manoeuvres including hesitation roll, lag back-cross manoeuvre and while checking position during formation flying. Some overstress situations occurred during unscheduled aerobatics sequences; sometimes, the pilot was distracted as he came out of one manoeuvre and entered the next.

LSA in Air Traffic Controllers

There were a number of cases where the lack of situation awareness applied to both air traffic controller and air crew. Both air traffic control (ATC) personnel and aircrew allowed themselves to engage in incomplete and non-standard communication which contributed to several near misses. A number of mishaps were

due to ineffective communication between the ATC and the pilot. In one case the controller allowed himself to become occupied with communication problems with another aircraft, and passed incomplete information to an aircraft performing a simulated forced landing. In another case the ATC controller became engrossed with a squadrons of helicopters requesting take off information instead of handling a four plane formation returning to base for a VFR pitch and landing. As a result a near miss occurred.

Several human factors issues also entered into the decision making performance of a relatively inexperienced ATC controller. This controller was only VFR qualified and yet required a good knowledge of IFR procedures to carry out his duties safely and effectively. During local night mission training, the controller apparently did not foresee the potential danger in departing a light civilian aircraft on one runway while a F5 was in the circuit for landing on a nearby runway. The F5 pilot was preoccupied with his fuel state, and did not recognise nor acknowledge the instruction to overshoot. The instructions that were issued by the duty controller were non-standard and vague. The supervisor assigned to monitor the duty controller thought that the controller had the situation under control and did not feel supervisory input was necessary. It resulted in a near miss incident where the aircraft came within 200-300 ft. (60-90 m.) of one another.

Case Studies on Accidents Related to LSA

Case 1 Cat A

While responding to an aircraft malfunction (drop in hydraulic pressure), the pilot, who was the sole occupant of the aircraft, failed to monitor the aircraft's descent rate and altitude. Apparently, the pilot chose to release the aircraft controls while the aircraft was in a turn at relatively low altitude in order to reset the hydraulic pressure. No ejection was attempted and no radio calls or emergency squawks were observed. The pilot sustained fatal injuries on initial impact. As a result of this accident, the validity of including a circuit breaker reset action item in the "loss of hydraulic pressure" non-critical emergency was examined. It was discovered that the circuit breaker only serves to electrically protect the hydraulic gauge and will not aid in resolving zero hydraulic pressure situations; therefore, the reset was removed from the checklist response. The primary rule in dealing with minor emergencies is to maintain aircraft control first and attend to the aircraft malfunction or emergency secondarily. It appears that in this accident the pilot became so engrossed in a non-critical emergency procedure that he failed to maintain his overall situation awareness.

Case 2 Cat A

The mission was planned as a routine navigation training exercise at 500 ft. (150 m.) AGL. Approximately 20 minutes after take-off, a Belgian F16 on an intersecting low level route spotted the T33 on his right side at the same altitude. The F16 pulled up and commenced a high left turn to execute a simulated attack. The T33 sighted the F16 and started a 180 degree defensive turn to the left in accordance with existing procedures. After completion of the turn the T33 was observed to level the wings, hit tree tops of a small hill and caught fire. The aircraft then impacted into a small grassy field 1200 ft. (360 m.) later and was destroyed. There were no attempts to eject; the two crew members sustained fatal injuries. On the basis of available information, it is suspected that the pilots concentrated on monitoring the F16 and failed to monitor and clear their own flight path adequately. In this particular case, the mission undertaken was not overseen closely enough to ensure that the participating pilots had commensurate experience and training.

Case 3 Cat A

While a Chinook helicopter was in a turning manoeuvre, the rear rotor struck and cut a telephone pole that was doubling as a light standard while it was taxing to the fuel facilities. This resulted in a chronic rotor imbalance followed by rotor blades striking the fuselage causing an explosive fire. The aircraft flipped over and came to rest among the fuel tanks. Survivors managed to exit the ball of flames suffering various degrees of burns. In this case, the aircraft was in a turning manoeuvre near a known obstacle within the minimum turning radii of that aircraft. The unit failed to observe the 75 ft. (22.5 m.) obstruction clearance limits required when taxiing.

Case 4 Cat B

A student pilot was executing a low level 180 degree turning autorotation from 250 ft. (75 m.)

above ground level. Halfway through the turn he allowed the rotor RPM to increase. The instructor assisted in controlling the RPM by increasing collective pitch and with the student still at the controls, returned his attention inside the cockpit to monitor the rotor RPM. At this point the student increased bank and attitude. The instructor took control when excessive bank and close proximity to the ground became evident. He levelled the aircraft and landed hard with speed short of the autorotation area. The helicopter sustained Category B damage. In this case the instructor was concerned with keeping the rotor RPM within limits and while concentrating on the RPM gauge lost situation awareness and allowed the aircraft to arrive at a position from which a safe recovery was not possible. The student, while performing a low-level 180 degree autorotation from 250 ft. (75 m.), allowed the aircraft to overbank and develops a nose low attitude during the turn resulting in a rapid rate of descent.

Case 5 Cat B

Two F18 aircraft collided while conducting a re-positioning exercise on an authorised William Tell workup mission. One aircraft lost the RH vertical stabilizer while the tactical lead aircraft sustained damage to the LH portion of the fuselage between the cockpit area and the nose radome. It was found that the pilot focussed his attention on the radar display for too long, thereby neglecting to clear the aircraft flight path during the lead change. The pilot of the tactical lead, did not exercise sufficient control of his formation, nor did he monitor the position of the wing man or the chase aircraft during the intercept. This resulted in a low level of situation awareness and created a hazardous situation which subsequently resulted in a mid-air collision. The lead pilot, expecting that the wing man was on his right side, mistook the chase aircraft located in this position as the wing man and acknowledged the lead change by calling "visual". A "building block" approach in preparing for this type of mission was not implemented which led to a situation where the pilots were unsure of how to accomplish the lead change procedure which was known to be a critical portion of the profile.

Case 6 Cat C

During a 4 plane ground attack mission the lead aircraft was in a left banked turn over a small, glassy-surfaced lake; on rolling out on heading he felt two bumps under the aircraft. Subsequently the RH engine compressor stalled. After the stall was cleared there was a long flame from the tail pipe and rising engine temperature. The engine was shut down. He also discovered that the pitch damper was not functioning resulting in a heavy control situation at 200 kias. In this case the pilot was distracted and failed to monitor his nose position and allowed the aircraft to descend and contact the trees while rolling out of a hard turn at low level.

REMEDIAL ACTIONS

Situation awareness provides the capacity to function in an anticipatory rather than a reactive mode. Traditionally, pilot training has concentrated mainly on developing physical flying skills, and knowledge of aircraft systems and procedures. Pilots tend to learn airmanship and develop situation awareness on the job. Usually hard lessons (accidents) are learned sporadically and are not part of a structured program. The increasing flow of information available from inside and outside the cockpit must be coordinated and utilised by the flight crew to achieve and maintain SA. This could be accomplished by cockpit management training that includes a thorough review of the event chains that lead to accidents, including a discussion of how to identify and interrupt error chains.

Instruction in the following would be valuable: awareness of local high potential conflict areas, lookout technique, situation awareness through effective listening out techniques, anticipation and needs for prompt reactions, the "see and be seen" principle which requires more cockpit time devoted to lookout. Simulator training in complete cockpit resource management during execution of a mission would provide the pilot with strong flight context experience and would be better than simulating isolated failures. A structured program could show the pilots how to recognise those situations where SA is usually lost and provides techniques to deal with these situations. Simulator training allows us to safely recreate in-flight situations rarely encountered in everyday flying. This adds to our experience file without risking injury, death and destruction of the aircraft. Subsequently, we may draw upon this file to react correctly. Properly designed simulator training scenarios will allow flight crews ample opportunity to become proficient in the use of these principles. This applies in varying degrees to all pilots, regardless of their type of aircraft or style of operation.
Introduction of a type-specific cockpit resource

management program in each of the CF flying operations will be complementary to existing programs. Current training efforts could be greatly enhanced by incorporating training that focuses specifically on the development of pilot SA. Such instruction could be internalised meaningfully when it is coupled with experience that can be provided in simulators and actual aircraft. Research should be carried out to determine the possibility of establishing an aircrew awareness management program for single seat fighter aircraft.

Human performance failures are primarily attention failures, and the mechanism of directing attention is not well understood. SA is a complex process of perception and pattern matching limited by working memory and attention capacity. Mechanisms such as attention sharing and automated processing may serve to alleviate these limitations to some degree.

CONCLUSION

SA is critical to pilot performance and survival in all types of flying operations. It fluctuates throughout any mission in any aircraft type. Maintenance of SA is not only for instrument flight. Loss of situation awareness has been implicated in many close calls and accidents. A unified operational definition of situation awareness is necessary, and perhaps tactical situation awareness and spatial disorientation should be considered as separate entities. Furthermore, a third category of SA could encompass the rest of pilot-induced cause factors.

For the near term, the only practical approach is to improve situation awareness training for pilots and air crews. Research has revealed that innovative training programs can reduce aircrew errors associated with situation awareness and judgment. The attention levels in air crews can be raised and habit patterns developed to handle threats in the flight environment. We need to identify the role and significance of attention problems in loss of situation awareness. The impact of selected contributory factors on attention problems, and those factors impacting on the pilot's ability to maintain situation awareness needs to be described. To minimise loss of situation awareness we should investigate mission and flight planning techniques. The mission plan should include response actions for each of the human performance events assigned a high probability of occurrence. Real-time, man-in-the-loop simulation training of pilots in various flight scenarios would improve the maintenance of awareness of situation information in flight. Similar training could also improve the performance of air traffic controllers.

REFERENCES

1. Endsley M.R., "Design and evaluation for situation awareness enhancement", *in* "Proceedings of the Human Factors Society-32nd Annual meeting 1988", Santa Monica CA: The Human Factors Society, 1988, pp 97-101.

2. Fracker M.L., "A theory of situation assessment; implications for measuring situation awareness", *in* "Proceedings of the Human Factors Society 32nd annual meeting, 1988", Santa Monica, CA: The Human Factors Society, 1988, pp 102-106.

3. Beringer D.B., Hancock P.A., "Exploring situation awareness: a review and the effects of stress on rectilinear normalization", *in* "Proceedings of the Fifth International Symposium on Aviation Psychology", Apr 17-20, 1989, pp 646-651.

4. Cheung B., Money K., Wright H., Bateman W., "Spatial disorientation implicated accidents in Canadian Forces 1982-1992", Aviat Space Environm Med, 1995 (in press).

5. Bellenkes A., Bason R., Yacavone D.O., "Spatial disorientation in naval aviation mishaps: a review of Class A incidents from 1980 through 1989", Aviat Space Environ Med, *63*, 1992, pp 128-31.

6. Gillingham K.K., "The spatial disorientation problem in the United States Air Force", J of Vestibular Res, *2*, 1992, pp 297-306.

7. Lyons T.J., Ercoline W.R., Freeman J.E., Gillingham K.K., "Classification problems of U.S. Air Force spatial disorientation accidents, 1989-1991", Aviat Space Environ Med, *65*, 1994, pp 147-152.

8. Pike W.S., "Pilot disorientation as the most frequent cause of fatal, weather-related accidents in UK Civil and general aviation", J of Navigation, *45*, 2, 1992, pp 229-240.

9. Holland D., Freeman J., "Loss of situational awareness and spatial disorientation in the USAF 1980-89", abstract presented at the Aerospace

Medical Association annual meeting, Miami, Florida, May 1992.

10. McNaughton G.B. "Vision in spatial disorientation (SDO) and loss of situational awareness", Proceedings of the Third Symposium on Aviation Psychology, April 22-25, 1983.

11. Kuipers A., Kappers A., van Holten C.R., van Bergen J.H.W., Oosterveld W.J. , "Spatial disorientation incidents in the RNLAF F16 and F5 aircraft and suggestion for prevention", *in* "Situational awareness in aerospace operations", Neuilly-sur-Seine, France, NATO/AGARD CP 478, 1990.

12. Lyons T.J., Ercoline W.R., Freeman J.E., Gillingham K.K., "Classification of US Air Force spatial disorientation accidents: 1989-91", Aviat Space Environ Med, *65*, 1994, pp 147-52.

13. Waag W.L., Houck M.R., "Tools for assessing situational awareness in an operational fighter environment", Aviat Space Environ Med, *65* , (5, suppl.), 1994, p A13-9.

14. Navathe P.D., Singh B., "An operational definition for spatial disorientation", Aviat Space Environ Med, *65*, 1994, pp 1153-5.

15. Department of National Defence Canada, "Flight Safety for the Canadian Forces", Canada, AGA-135-001/AA-001, 1988.

LOSS OF AVIATION SITUATIONAL AWARENESS CAUSED BY A BIRD-AIRCRAFT COLLISION

by

José L. García-Alcón MD.PhD. LtCol MC SAF
Chief Flight Surgeon
and
Juan M. Moreno-Vázquez MD.PhD.
Senior Professor of Physiology
23th Wing of Spanish Air Force
Talavera AFB
Badajoz- Spain

INTRODUCTION.

The history of aerial combat shows that tactical mission effectiveness depends on a few of the superior fighter-attack pilots. These few superior pilots appear to possess a heightened situational awareness that is decisive in the complex and highly fluid air combat environment.

Pilot situational awareness is not a well-understood concept. Most attempts at understanding situation awareness have relied almost exclusively on subjective reports, and have not led to a clear understanding of the concept. One approach to lending structure to defining situational awareness measures is to decide the information elements the pilot uses to develop an adequate "state of knowing" concerning to specific tasks, and to optimize the mission performance.

Operational requirements for situational awareness have implications for flight safety and combat effectiveness. After a series of interviews with air combat experts, some authors developed a taxonomy of skills, traits and performance measures important for air-to-air combat which included aggressiveness, decisiveness, hands-on flying skills, knowledge, ability and situational awareness. This last concept was described as "probably the sum of many perceptual and cognitive skills". (1)

In 1984, loss of situational awareness, was cited as a probable contributory factor in twenty out of forty one USAF operator-factor accident review. (2) And loss of situational awareness is related to and a potential contributor to spatial disorientation. However, situational awareness is intended to be a more wide term, encompassing more than spatial disorientation references, and include more clearly psychological aspects of attention and cognition as well as sensory physiology considerations.

In addition, future combat aircraft should be capable of flying anywhere, anytime and to engage air-to-air or air-to-ground attacks against a very unfavourable ratio of enemy forces. It is a clear consequence that the pilot's role will ever more turning to supervisory control, handling engines and the managing short term tactical decisions. In this way, and because the airplanes are of comparable technology in various countries, the quality of situational awareness will probably explain the within pilot's variation of performances. (3)

Thus, the situational awareness, leads to ability to detect the geographical position and to understand where one comes from and where one goes to. The mental process involved in this activity follows a dynamic model and encompassing ability to position oneself, spatial perception with emphasis on vision and vestibular interactions etc. and after logical feedback, the pilot will obtain a satisfactory situation awareness that contributes to enhance the pilot's capabilities and performance, and diminishing the occurrence of spatial disorientation.

Some authors postulate that at least six generic skills are essential to keep and enhance the situational awareness: (4)

One.- Height sensitivity to extremely short-duration, low-intensity cues in the external stimulus field.

Two.- Early acquisition of situation-determinant.

Three.- Rapid integration between objectives and situational characteristics.

Four.- High-speed automated processing of acquired information under conditions of time-urgency and stress.

Five.- Virtually instantaneous situation assessment from minimal input information. Finally

Six.- Direct apprehension of situation dynamics and trends

The first three skills listed have been named as the near-threshold skill, and are involved in acquiring, analysing, processing and producing responses to sensory stimuli near, at, or below the level of conscious awareness. It can be developed and enhanced through special training. If these skill can be systematically and efficiently trained, we suppose that all pilots have the sufficient potential to enhance combat capability and flying safety, unconstrained by aircraft type, mission requirements or operational environments.

Thus, the failure of whichever of these skills, can lead to loss of situational awareness. In addition, aviation accidents arenot produced by a fact alone. On the contrary, most of the times, an accident is the final consequence of a chain of incidents. Loss of situational awareness can be the first step in that chain or, sometimes, the last and more decisive fact that cause the accident.

In this way, we report a case, which the loss of situational awareness was the secondary consequence of an incident, and was on the point of provoking a fatal accident itself.

CASE DESCRIPTION

A Spanish F-5B aircraft was performing a laser target designation mission last April, flying over an area located in the southwest of Spain. Its altitude was three thousand feet and its airspeed was four hundred knots approximately. The aircrew was formed by two pilots. The

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

first pilot, seated in the forward cockpit, was twenty-six years old. He had flown about one thousand and one hundred hours in the F-5 aircraft. The second pilot and laser operator, seated in the rear cockpit, was twenty-eight years old with a flight experience in this airplane of one thousand hours. Both pilots were captains and their flight qualification was combat ready (CR), according to NATO and Spanish regulations. They performed three to four laser target designation missions every week.

The flight was being developed normally but as they were approaching the target area, the aircraft suffered suddenly a great impact on the left side of the front cockpit. The cause of this huge impact was initially unknown. After a few seconds the first pilot realized that the object that had collided against the canopy was a big bird. Later the flight safety officer of the Wing identified the feathers remaining inside both cockpits, as belonging to a vulture.

Neither one had observed any bird presence previously during the flight time before the accident. However, during spring time, the southwest of Spain is a nesting area for several kinds of birds, specially big birds as storks, vultures, great bustards, etc. Each specie has its own flight envelope, for example, the white stork flies usually between three hundred and five hundred feet; the great bustard flies normally around two hundred feet, and the Spanish vulture can climb as high as ten thousand feet or higher.

The Spanish vulture is a big bird with a mean span of two meters and a mean weight of sixteen kilos. The vulture has not any enemies and its powerful flight and great size are factors that contribute to an excess of self-confidence. Consequently, vultures never attempt to elude the collision with another flying object.

Some instants before the impact, the first pilot was slightly inclined to the instrument's panel, changing the radio frequency. In this position, his head was under the frame of the cockpit. Because of that, his head was partially protected. The laser operator, in the rear seat, was preparing the laser target designation device located in the left side of the rear cockpit. The visor of his helmet was up to help the vision through the laser telescope.

The bird collided against the upper left side of the windshield and against the left side of the front canopy simultaneously. Both glasses were broken out. The bird and glass fragments of the canopy and windshield went inside the forward cockpit and hit the first pilot's helmet laterally.

He was slightly injured by small fragments that hit his face and hands. His head was protected by the oxygen mask and helmet and his hands were protected with gloves.

The ensemble composed by glass fragments and parts of the bird's body went all the way through the rear cockpit, breaking the middle panel that separates both cockpits. Then, it hit the left hand and the left side of the second pilot's head. The ensemble continued its way through the rear cockpit, leaving it through the left rear side of the canopy, hitting later the tail of the airplane.

At the moment of the impact, the first pilot lost the situational awareness, because his helmet visor was covered by a lay of the bird's blood. This lay of blood made the vision impossible and, consequently, the aircraft control. So, and to be allowed to clean his helmet visor, the first pilot gave the aircraft control to the second pilot quickly, ordering "you have it" through the intercom system.

The first pilot believed that the only damage in the airplane was the front cockpit glass, and was ignorant of two important aspects of the situation: first, the second pilot was severely injured, and second, the intercom system was disabled.

The first pilot was surprised when the second pilot did not answer to his order by the normal procedure, having to say "I have it" and by the fact he felt how the airplane started to descend. Because of all that, the aircraft was uncontrolled during an indeterminate time. In those few seconds, the first pilot lost of situational awareness, and when the first pilot took the control again, the aircraft had descended from three thousand feet to two hundred feet.

After the aircraft control was recovered, the pilot established a course heading to the Air Base and he transmitted his situation, making a report of damages, since the external communication system (UHF communication), remained in order. During the return flight, the first pilot attempted unsuccessfully to talk with the second pilot several times. He knew that the second pilot was alive, though injured, because he could see his head movements by the rear-view mirror. Afterwards, the flight surgeon verified, that the second pilot was conscious all time, but slightly shocked because of the head trauma.

After landing, the aircrew was helped by the rescue team, and carried to a medical centre.

The first pilot was unharmed.

The injuries of the second pilot were:

-Left eye enucleation, with small fractures of the upper limb of his orbit.

-Several face erosions.

-Comminute fracture of the first metacarpal bone of his left hand.

-Double fracture of the second metacarpal bone of his left hand too.

The treatment at the hospital spent fifteen days and the pilot requires two surgical operations. After six months he was permanently disqualified for flight duties.

The report made by engineers, about the total damages caused in the aircraft was as follows

-Front windshield destroyed and windshield frame deformed.

-Front canopy glass destroyed.

-Front jettison canopy tube cut.

-The intercockpit bullhead destroyed.

-The Connection links of rear canopy drive mechanism bent.

-Some Structural damages in the drive mechanism of the rear canopy that remain unavailable.

-Scissions at the leading edge of the vertical stabilizer.

-Damages in the coating and honeycomb of the vertical stabilizer.

-Pilot's intercom system destroyed.

DISCUSSION.

Before starting any discussion, the authors point out that this is not a conventional paper but an ensemble of considerations about the loss of situational awareness, its consequences and the way to avoid it.

Unquestionably, in the case mentioned above, the pilot suddenly suffered a loss of situational awareness originated by an exterior cause, as it was the collision with a vulture. As the following research revealed, performed by the Flight Safety Officer, the loss of the situational awareness consisted of:

A.- The loss of the first pilot's vision, due to the blood bird's lay on his helmet.

B.- The loss of communications between both pilots.

In reality, the loss of the situational awareness was caused in part by the first pilot, because immediately after the impact, from which he had not been affected, his

situational awareness was normal. He had exterior visual information, and radio communication; he was not hurt, and he kept the control of the airplane.

Loss of situation awareness, may begin when a limited supply of attention is distributed among several elements of a situation. (5). Because attention is limited, the person may allocate more attention to some elements than others depending upon the priority he assigns to each. Priorities, in turn, should be decided by the degree to which each element threatens or contributes to successful task completion. In our case, the first pilot made the mistake of assuming the impact had affected only his cockpit and he transferred the control of the aircraft to the second pilot, dedicating his attention to cleaning the helmet visor in order to recover a perfect vision. As the second pilot could not take the control of the aircraft, since he was hurt, the aircraft remained without any control for a minimum fraction of time. From our point of view, this fact is a wrong priorities' assignation. In this case the pilot should be assigning the priority to the airplane control.

It is important to remember that the first pilot quickly recovered the situational awareness and, due to his experience and training, he was able to recognize the dangerous situation, and resolve the problem immediately.

There is not doubt that if the aircraft had been a single-seat F-5, the loss of situational awareness would have not existed, since the pilot has not any possibility to transfer the control of the airplane to anybody and thus, he would have only raised the visor of his helmet, without trying to clean it.

Nevertheless, we estimate that pilot has to make a great effort to keep himself always perceptive to the external environment, where the information for flying comes from. The aircrew must be trained (even with psychological support) to perform an adequate distribution of attention priorities in each flight phase and its circumstances. Slight distractions, or different allocation of a limited supply of attention, can lead to a loss of the situational awareness, partially or completely, and can affect critically the Flight Safety during periods of peace or can affect the mission success at war.

In conclusion, we estimate that experience and training are fundamental factors to reduce the possibility of losing the situational awareness, and perhaps, the best way to training the aircrew in to get, to keep and to enhance the situational awareness can be the aircraft simulator. Further more, in case of loss of the situational awareness, a trained and experienced pilot has more chances for a more rapid recovery of it.

REFERENCES.

1.- Kelly, MJ. Performance measurement during simulated Air-to-air combat. Human factors. 1.988 30(4):495-506

2.- Pruden, AL. Loss of aircraft attitude awareness: Impact on USAF. New Techonology - Old Problems. 1985 In: Aircraft attitude Awareness Workshop Proceedings. McNaughton, GB (Ed) Flight Diynamics Laboratory. Wright Patterson Air Force Base. Ohio. pp 1-2-1 to 6.

3.- Menu JP, Amalberti R. Les déterminants de l'appréciation de la situation tactique et le développement de systèmes d'aides ergonomiques. In Situational Awareness in Aerospace Operations. AGARD/AMP CP n°478. 6-1. 1.989

4.- Secrist GE and al. Situational awareness the trainability of the near-threshold information acquisition dimension. Aviat.Space & Env. Med. 64 885-892, 1993

5.- Fracker ML. Attention gradients in situation awareness. In Situational Awareness in Aerospace Operations. AGARD/AMP CP n°478. 6-1. 1.989

Determinants of Situational Awareness in U. S. Air Force F-15 Pilots

Thomas R. Carretta Malcolm James Ree
Armstrong Laboratory Human Resources Directorate
7909 Lindbergh Drive
Brooks Air Force Base, TX 78235-5352
United States

SUMMARY

Situational awareness (SA) is often cited as essential in pilot performance in both military and civil aviation. A study was conducted to investigate whether SA in U. S. Air Force (USAF) F-15 pilots could be predicted using the constructs of general cognitive ability, psychomotor ability, personality, and previous job (i.e., flying) experience. These constructs were chosen because they have demonstrated validity for the prediction of performance in a wide variety of military and civilian jobs, including pilots. The participants were 171 active duty F-15 A/C pilots. Test scores, flying experience data, and supervisor and peer ratings of SA were collected at the pilots' duty locations. The first unrotated principal component extracted from the supervisor and peer ratings accounted for 92.5 percent of the variability of the ratings, which indicated substantial agreement between supervisors and peers. The unrotated first principal component was used as the measure of job performance (i.e., SA). Flying experience as measured by number of hours in the F-15 was the best predictor of SA. After controlling for the effects of F-15 experience, the measure of general cognitive ability based in divided attention, spatial reasoning, and working memory was predictive of SA. Psychomotor and personality measures did not add to the predictiveness of general cognitive ability. With additional F-15 flying experience, it is expected that pilots would improve their SA ratings. Implications for pilot selection and follow-on research are discussed.

1. INTRODUCTION

Poor flying performance resulting in a failed mission and/or a loss of aircraft and life is often blamed on a lack of situational awareness (SA). The ability of a pilot to know location in space and time, and to keep track of other aspects of the dynamic environment of flight, are the common elements of definitions of SA.

The impetus for this study was provided by the U. S. Air Force (USAF) Chief of Staff. He directed USAF research laboratories to determine which attributes enabled a pilot to develop and maintain SA, especially in the high performance F-15 jet aircraft. The definition of SA adopted for this study was:

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

"A pilot's continuous perception of self
and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception."

Thus, SA was seen as a complex construct involving perception, processing and interpreting data, forecasting, and behavior. A failure at any one of these steps would result in poor SA and less than optimal performance.

Much of the published literature regarding SA in pilots suffers from a lack of adequate sample size and therefore, a lack of statistical power for detecting significant effects. A review of seven recent studies of SA (1, 2, 3, 4, 5, 6, 7) showed sample sizes that ranged from 8 to 56 with an average of 21.75 participants. The likelihood of detecting significant correlational effects is quite low with such small samples. Larger samples are clearly needed in order to improve the likelihood of detecting significant effects.

Several other methodological problems sometimes occur that effect our ability to interpret results from studies of pilot ability. These include (a) the failure to examine the construct validity of measures used in validity studies, (b) the failure to correct for level of job experience when examining the relationship between ability and job performance, and (c) methodological problems that occur as a result of using range restricted samples and measures with unknown reliability.

This concurrent validation study (correlation design) was conducted from a personnel selection standpoint. The purpose was to determine which human attributes were predictive of SA in the F-15 A/C as judged by peers and supervisors. The F-15 A/C was chosen in order to investigate SA in the air-to-air mission, as it is the premier intercept aircraft for the USAF.

2. METHOD

Participants

The participants were 171 active duty USAF F-15 A/C pilots. The pilots were tested at their duty locations in Eglin, FL, Elmindorf, AK, Kadena, Japan, and Langley, VA. They were all male college graduates ranging in age from about 24 to 45 years and ranging in military rank from first lieutenant to lieutenant colonel. They had between 1 and 20 years post-pilot-training flying experience and between 88 and 2,007 F-15 flying hours and between 193 and 2,805 total post-pilot-training flying hours.

Measures

Predictors. Both multiple empirical studies and meta-analyses have identified three predictors that are valid for almost all job performance criteria. These are general cognitive ability (8, 9, 10, 11), psychomotor skill (8), and the personality construct of "conscientiousness" (12). The USAF currently collects measures of general cognitive ability and psychomotor ability for pilot training applicants (13) in the Air Force Officer Qualifying Test

(AFOQT) and the Basic Attributes Test (BAT).

SA predictors included several measures of general cognitive ability, psychomotor skills, and personality. General cognitive ability was measured with cognitive components tests (14) of processing speed, near threshold processing, reasoning, velocity estimation, and working memory. Content for these measures included verbal, quantitative, and spatial. Psychomotor measures included tests of multilimb coordination, aiming, control precision, reaction time, and rate control. Personality was assessed using a measure of the Big 5 (15) constructs of extroversion, emotional stability, agreeableness, conscientiousness, and openness. More detailed descriptions of the specific tests appear in Carretta, Perry, and Ree (16). Cognitive, psychomotor, and personality measures were administered by a computer-based system (13).

Criteria. The criterion (17) was derived from multiple supervisory and peer SA ratings developed from task analyses with seven experienced F-15 pilots that served as subject matter experts (SMEs). Each SME had over 1,000 fighter aircraft hours. These fighter pilot SMEs identified tasks essential to air combat success and required for SA. This resulted in 31 behavioral items representing personal traits and job tasks related to SA.

Supervisor rating items represent general traits, tactical game plan, systems operation, communication, information interpretation, and tactical employment. Standardized definitions for each of the items were provided to every rater to establish consistency. Each of the 31 items was rated on a six point Likert scale from 1- "Acceptable" to 6- "Outstanding." Supervisors also provided ratings on overall fighter ability and SA ability on the same six point scale. All of the pilots that participated in the study were rated by multiple raters. This rating scale is provided in Carretta, Perry, and Ree (16).

For peer ratings, pilots rated other pilots in their squadron with whom they had flown. Overall fighter ability and SA ability were scored on a six point Likert scale ranging from 1- "Acceptable" to 6- "Outstanding." Pilots also rank-ordered peers from 1- "the best I've flown with " to N, “the number of peers rated”, indicating their judged standing on the trait of SA.

Control variables. The consequences of job experience are important in understanding the relationship between ability and performance (18). The American Psychological Association (19) in it’s *Principles for the Validation and Use of Personnel Selection Procedures* states that validation studies using job incumbents should control for the effects of maturity, increased job knowledge, and motivation. This was accomplished by statistically holding flying experience constant.

Procedure

Participants were tested on the computerized battery at their operational air bases. Supervisory and peer ratings of SA were collected independently. The cognitive test scores that included both accuracy and time were formed into ratios to yield a measure of correct responses per unit time (14).

Analyses

The issue of single versus multiple criteria was addressed by investigation of the unrotated principal components of the SA ratings. If the first unrotated principal component accounted for a large portion of the variance and the succeeding components a small proportion, a single (or composite) criterion would be preferable. This single, composite criterion would also be more reliable than the individual rating scales.

The job (flying) experience control variables were selected through regression analyses. Regressions of the criterion on F-15 hours, F-15 hours squared, F-15 hours cubed, total flying hours, total flying hours squared, and total flying hours cubed were computed. The squared and cubed terms were necessary to account for any non-linear relationships. Only those variables that contributed significantly to the regression were kept. Validity of the tests to predict SA was assessed with the effect of flying experience controlled for statistically by partial correlation and by entering flying experience control variables into the regression equations with test scores. Test scores were included in the regression equations if they showed a significant partial correlation with the criterion.

The F-15 pilots who participated in this study had undergone extensive selection, screening, and training that reduced the variability of the attributes measured by the test scores and SA ratings. They had completed college, applied for and been selected for officer commissioning and pilot training, and graduated with high class standing from pilot training. They clearly represent the best of the pilot training applicants. Due to restriction in range, the correlations for the predictors were downwardly biased estimates of the population values (20, 21, 22).

To understand the relationship among the predictors, principal components and factor analyses were conducted on all sets of predictors that showed a significant partial correlation (flying experience being held constant) with the criterion. The two predictor sets from which composites were formed included the cognitive tests and the psychomotor tests and were developed through partial correlation analyses. The personality scales were not factor analyzed because they were already the consequence of factor theory.

Tests of linear regression models were used to assess the validity and incremental validity of the predictors. The first linear model (M1) was a full model that included flying experience control variables, a general cognitive ability composite (GCA), a psychomotor composite (PM), and a measure of conscientiousness (CON). The personality construct of CON was included because past research (12) suggested it would be predictive. M1 also included interactions among the cognitive, psychomotor, and personality composites.

The first reduced linear model (M2) was M1 with the three-way interaction removed and similarly, M3 was M1 with all interactions removed. Reduced models M4, M5, and M6 contained the flying experience control variables and combinations of two of the three

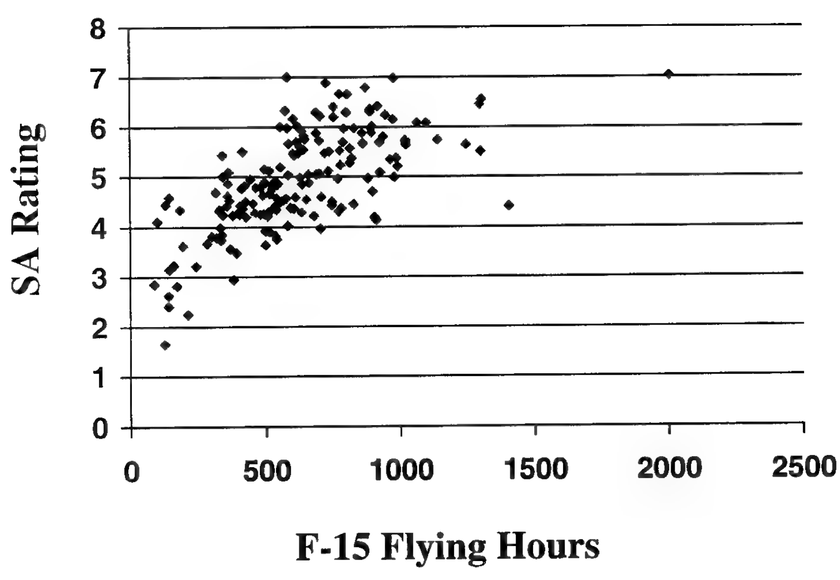
predictors found in model M3. Models M7, M8, and M9 contained the control variables and one of the predictor composites. M10 contained the control variables only. M7, M8, and M9 were tested against M10.

The statistical testing began by removing the interaction terms and then the individual GCA, PM, and CON composites to determine which were statistically significant predictors of SA. All statistical tests (i.e., the identification of the flying experience control variables, the partial correlations, and the linear models analyses) were conducted using a $p \leq .05$ Type I error rate.

3. RESULTS

Principal components analysis of SA ratings yielded only one eigenvalue greater than one. This factor accounted for 92.5% of the variance in the ratings, demonstrating that a single criterion was preferable to multiple criteria. Consequently, the first unrotated principal component of the ratings was the SA criterion.

The two flying experience control variables found useful were F-15 flying hours and F-15 flying hours squared (R = .704). Generally, as the number of F-15 hours increased, so did the SA ratings. Figure 1 shows a scatterplot of the first principal component of the SA ratings by total number of F-15 flying hours. The SA ratings were rescaled for Figure 1 to have a mean of 5 and a standard deviation of 1.



<u>Figure 1</u>. **SA Ratings by F-15 Flying Hours**

The matrix of partial correlations holding flying experience constant disclosed six tests (four cognitive and two psychomotor) as significant predictors of SA. These were measures of verbal working memory, spatial reasoning, divided attention, spatial working memory, aiming, and attention, reaction time, and rate control. Principal components analysis of the four cognitive tests that had significant partial correlations showed that the first component accounted for 51% of the variance. The unit-weighted sum of these tests became the measure of GCA.

A principal components analysis of the two psychomotor tests was not conducted as at least three variables are required. The unit-weighted composite of the two tests became the PM composite.

As the Big 5 personality measure was already based on factor analyses, no principal components analysis was done.

Table 1.
Linear Models Tested

Model	Predictor Variables	R	Models Tested	ΔR
M1	F-15 hours, F-15 hours squared, GCA, PM, CON, (GCAxPMxCON), (GCAxPM), (GCAxCON), (PMxCON)	.741		
M2	F-15 hours, F-15 hours squared, GCA, PM, CON, (GCAxPM), (GCAxCON), (PMxCON)	.733	1 vs. 2	.008
M3	F-15 hours, F-15 hours squared, GCA, PM, CON	.732	2 vs. 3	.001
M4	F-15 hours, F-15 hours squared, GCA, PM	.732	3 vs. 4	.000
M5	F-15 hours, F-15 hours squared, GCA, CON	.727	3 vs. 5	.005
M6	F-15 hours, F-15 hours squared, PM, CON	.713	3 vs. 6	.019
M7	F-15 hours, F-15 hours squared, GCA	.727	4 vs. 7	.005
			5 vs. 7	.000
M8	F-15 hours, F-15 hours squared, PM	.712	4 vs. 8	.020*
M9	F-15 hours, F-15 hours squared, CON	.704	5 vs. 9	.023*
M10	F-15 hours, F-15 hours squared	.704	7 vs. 10	.023*
			8 vs. 10	.008
			9 vs. 10	.000

Note. ΔR is the difference in multiple correlations between the two models.
* $p < .01$

The sum of the items for the conscientiousness scale became the CON composite.

All linear regression models were significantly correlated with the criterion (see Table 1). Statistical tests showed that the interactions were not significant. This was determined by testing M2 against M1 and testing M3 against M2. To test the incremental validity of each type of predictor (GCA, PM, CON), M4, M5, and M6 were tested against M3. No differences were found for M3 versus M4 or M3 versus M5. However, when GCA was removed, M3 versus M6, a significant difference was found. Further, M4 versus M7 and M5 versus M7 were tested and found not to differ. This indicated that PM and CON were not incremental to GCA for prediction of SA. Also, the comparisons for M7, M8, and M9 versus M10, revealed that only GCA provided incremental validity beyond F-15 flying experience.

4. DISCUSSION

Flying experience in the F-15, which brings F-15 job knowledge, was the most predictive variable. This is consistent with Hunter's (23) demonstration that ability influences job performance via the accumulation of job knowledge. The implication is that if pilots were allowed to acquire more flying hours, their job knowledge would be expected to increase as would their SA.

When job experience was held constant in the regressions, general cognitive ability was found to be predictive of the criterion. The psychomotor score and the personality trait of conscientiousness were not.

The results for general cognitive ability are in agreement with several recent studies (9, 10, 24, 25). McHenry et al. (1990) demonstrated that general cognitive ability was predictive of job performance in nine jobs. Olea and Ree (10) showed the predictive utility of general cognitive ability for several pilot and navigator training criteria including academic grades and hands-on flying performance work samples.

Ree, Carretta, and Teachout (25) in a causal model analysis, demonstrated that general cognitive ability led to the acquisition of flying job knowledge, both prior to and during training. Further, they found that general cognitive ability worked through job knowledge acquisition to influence hands-on flying performance during primary and advanced jet flying training. Ree and Carretta (24) provide a broader discussion of the role of general cognitive ability in pilot selection.

That the psychomotor composite was not predictive of SA when job experience was held constant was contrary to Carretta and Ree (13). They demonstrated the incremental validity of psychomotor measures for predicting pilot training performance. Ree and Carretta (26) have demonstrated that psychomotor tests measure general cognitive ability, along with general and specific psychomotor ability. It is likely that the constant training, provided by frequently flying the F-15 aircraft, served to reduce to almost vanishing, individual differences in general and specific

psychomotor ability. This would account for their lack of validity for this criterion.

In contrast to the findings of Tett, et al. (12), conscientiousness failed to be a significant predictor of the criterion. The reasons for this failure cannot be found in these data.

The implications of the study are straightforward. The first implication is based on the finding that a greater number of F-15 flying hours was related to higher ratings of SA. The more hours pilots are permitted to spend in the F-15 cockpit, the better their SA can be predicted to be. The second implication is related to personnel selection. Current US Air Force pilot candidate selection procedures (e.g., AFOQT, BAT) rely heavily on measures of the construct found to be predictive of SA: general cognitive ability. Future revisions of pilot selection instruments should retain measures of general cognitive ability.

There are several issues to be addressed in the measurement of cognitive ability. It is necessary to improve the accuracy and completeness of our measures as found in the AFOQT and BAT. The measurement of cognitive ability using several different contents implies that different test level traits may be used. These are often referred to as first-order factors or constructs. Many are familiar, such as verbal and quantitative, and some have emerged more recently from models of cognitive components, such as spatial working memory, verbal working memory, and spatial reasoning.

Although the equivalence of new cognitive components and first-order factors remains to be fully investigated, it may be that new cognitive components offer measurement of cognitive ability with almost no content in the usual sense. That is, the new cognitive component tests frequently do not require previous learning other than the language requirements of the instructions. We speculate that problems of adverse impact on minority groups and women might be reduced or avoided if new cognitive tests with little or no learned content are added to the pilot selection system (27, 28).

Follow-on USAF research is focused on activities needed to transition these experimental SA measures to an advanced technology development (and eventually an operational) phase. Studies are planned to examine the psychometric characteristics of these tests including reliability, validity against training performance criteria, relationship to other selection instruments, retest gains, and subgroup differences (i.e., sex and ethnicity) in test performance.

REFERENCES

1. Andre, A. D., Wickens, C. D., Moorman, L., & Boschelli, M. M. (1991). Display formatting techniques for improving situation awareness in the aircraft cockpit. *The International Journal of Aviation Psychology*, *1*, 205-218.

2. Bolstad, C. A. (1991). Individual pilot differences related to situation awareness. *Proceedings of the Human Factors Society, 35th annual meeting* (pp. 52-56), San Francisco, CA.

3. Endsley, M. R. (1990). Predictive

utility of an objective measure of situation awareness. *Proceedings of the Human Factors Society, 34th annual meeting* (pp. 41-45), Orlando, FL.

4. Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology*, *4*, 241-264.

5. Fracker, M. L. (1991). *Measures of situation awareness: An experimentalevaluation* (AL-TR-1991-0127). Wright-Patterson AFB, OH: Human Engineering Division, Crew Systems Directorate.

6. Fracker, M. L., & Davis, S. A. (1991). *Explicit, implicit, and subjective rating measures of situation awareness in a monitoring task* (AL-TR-1991-0091). Wright-Patterson AFB, OH: Human Engineering Division, Crew Systems Directorate.

7. Kass, S. J., Herschler, D. A., & Companion, M. A. (1991). Training situational awareness through pattern recognition in a battlefield environment. *Military Psychology*, *3*, 105-112.

8. Hunter, J .E., & Hunter, R. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96*, 72-98.

9. McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

10. Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology*, *79*, 845-851.

11. Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1*, 86-89.

12. Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

13. Carretta, T. R., & Ree, M. J. (1994). Pilot Candidate Selection Method (PCSM): Sources of validity. *The International Journal of Aviation Psychology, 4*, 103-117.

14. Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology, 120*, 375-405.

15. Goldberg, L. R. (1990). An alternate "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

16. Carretta, T. R., Perry, D. C., Jr., & Ree, M. J. (in press). Prediction of situational awareness in F-15 pilots. *The International Journal of Aviation Psychology.*

17. Houck, M. R., Whitaker, L. A., & Kendall, R. R. (1991). *Behavioral taxonomy for air combat: F-15 defensive counter-air mission* (UDR-TR-91-147). Dayton, OH: University of Dayton Research Institute.

18. Schmidt, F. L., Hunter, J. E., & Outerbridge, A .N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.

19. American Psychological Association Division of Industrial-Organizational Psychology (1980). *Principles for the validation and use of personnel selection procedures*, (2nd ed.). Berkeley, CA: Author.

20. Thorndike, R. L. (1949). *Personnel selection.* New York: Wiley.

21. Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62* (Section A, Part 1) 28-30.

22. Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology, 79*, 298-301.

23. Hunter, J. E. (1983). A causal analysis of cognitive ability, job performance, and supervisory ratings. In F. Landy, S. Zedeck, and J. Cleveland, *Performance measurement and theory*, Hillsdale, NJ: Erlbaum.

24. Ree, M. J., & Carretta, T R. (in press). The central role of *g* in military pilot selection. *The International Journal of Aviation Psychology.*

25. Ree, M. J., Carretta, T. R., & Teachout, M. S. (in press). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology.*

26. Ree, M. J., & Carretta, T. R. (1994). The correlation of general cognitive ability and psychomotor tracking tests. *International Journal of Selection and Assessment, 2*, 209-216.

27. Carretta, T. R., & Ree, M. J. (in press). Near identity of cognitive structures in sex and ethnic groups. *Personality and Individual Differences.*

28. Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement, 55*, 268-277.

Near-Threshold Visual Perception and Manual Attitude Tracking: Dual-Task Performance and Implications for Situational Awareness

Jeremy M. A. Beer, Robert A. Gallaway
Systems Research Laboratories, Inc., Brooks AFB, TX 78235 USA
Fred H. Previc
U.S.A.F. Armstrong Laboratory, Brooks AFB, TX 78235 USA

SUMMARY
This study tested viewers' near-threshold recognition throughout the visual field, examined the effect of this task on a concurrent tracking task, and tested whether individuals' performance on each cockpit task would predict dual-task performance. An underlying assumption was that efficient multitasking supports SA. Experiment 1 measured recognition duration thresholds at different nonfoveal locations. Subjects classified briefly-presented aircraft as fighters or nonfighters. An adaptive procedure adjusted viewing duration to reach 75% performance. Critical tracking ability was assessed also, using a central attitude display. Recognition deteriorated with eccentricity, and a wide distribution of threshold-recognition and tracking abilities was found. Experiment 2 combined the two tasks; instructions emphasized recognition as primary. Measures included decrease in recognition accuracy, change in response latency, and increase in tracking RMS in dual-task conditions. Thresholds from Experiment 1 predicted dual-task recognition (viewers requiring longer durations classified fewer aircraft correctly), but not the increase in tracking error. Under dual-task stress, viewers with low recognition thresholds were less likely to abandon that primary task early in favor of tracking. Notably, critical tracking ability was linked to success preserving aircraft recognition: Viewers who tolerated higher instability before crashing suffered less in dual-task recognition. Field biases were identified in near-threshold performance under workload; this parallels visual search findings and is consistent with a spatially biased attention system. Findings are potentially relevant to SA assessment/selection and to the design of cockpit displays.

Introduction

The observer's ability to interpret near-threshold visual stimuli, images that are near sensory threshold levels for duration or contrast, has been proposed as a foundation for situational awareness. Hartman and Secrist [1] quoted an evocative description of a pilot who "can extract more from a faint tangle of condensation trails, or a distant flickering dot, than he has any reason or right to do." According to this principle, pilots who can process dim or fleeting targets most efficiently would be at an advantage when global situation perception is required.

Several questions must be answered before this principle is established as a useful aviation construct. First, real-world targets, including aircraft, can emerge not just where the pilot happens to be looking but anywhere in the visual field, and it has not yet been determined how near-threshold recognition of such targets varies throughout the visual field. In addition, it has not yet been demonstrated conclusively that near-threshold performance predicts the viewer's ability to maintain a global situational percept.

This project begins to address these questions. An underlying assumption, supported in the aviation literature [2, 3, 4], is that a requirement for maintaining SA is the ability to perform several perceptual and cognitive tasks effectively at once. Good performance on two concurrent visual tasks, therefore, would be interpreted as evidence that the viewer possesses skills that support highly-developed SA. Two experiments used complementary measures to test the distribution of near-threshold performance throughout the visual field, examined the attention drain that near-threshold recognition would place on a simultaneous manual attitude task, and tested whether individuals' performance on each of these tasks would predict their ability to perform both together.

Experiment 1: Recognition thresholds and manual tracking performance, single-task

Experiment 1 tested the spatial distribution of threshold recognition performance, using an adaptive threshold estimation method. In addition, Experiment 1 measured each viewer's

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

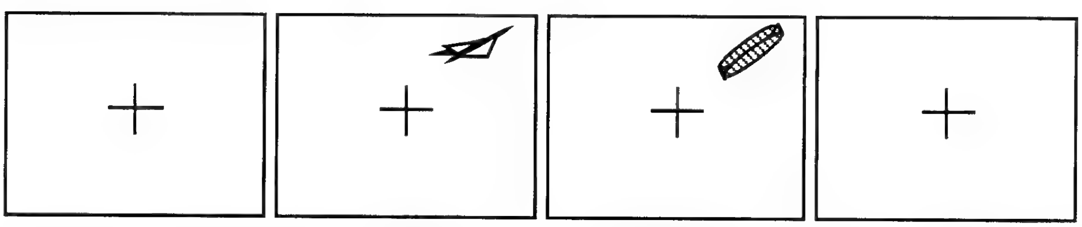
overall threshold recognition performance, as well as his or her manual tracking ability.

Moving stimuli away from the fixation location is similar to decreasing contrast or viewing duration in its perceptual consequences, because it reduces available sensory information. In addition, performance in the periphery can be subject to visual field asymmetries [5, 6]. These effects are important because flight tasks place considerable demands on peripheral processing; scanning a scene or an instrument panel in a succession of glances depends on processing features in the periphery [7, p. 159]. Duration thresholds were measured for recognition at three retinal eccentricities in the four visual field quadrants, to determine how performance would fall off in the periphery. Eighteen right-handed and two left-handed Armstrong Laboratory employees participated, completing four training sessions and one test session. Considerable training was included in order to avoid testing subjects who were still climbing a steep portion of the learning curve.

Stimulus sequences were generated using an Iris workstation. A collimated display system was used, to simulate distance focus. A joystick was used to collect tracking data and two-alternative recognition responses. In the recognition task, aircraft were displayed at twelve possible visual locations. Viewers identified the aircraft as belonging to a modern fighter (e.g. F-16, Mirage) or a non-fighter (e.g. Boeing 747, Cessna 150) target set, which had been studied previously. Aircraft were shown from four viewing angles. Images were size-normalized to eliminate scale cues between differently sized planes. Images were generally 2 deg wide.

In each trial, a central fixation cross was displayed. An aircraft was then displayed briefly in the upper right, upper left, lower left, or lower right quadrant of the screen, at an eccentricity of 5, 9, or 13 deg (Figure 1). A mask was then displayed for 500 ms at the same location as the plane, after which the viewer pressed one of two buttons on the joystick to indicate which group the aircraft belonged to. Auditory feedback was provided during training.

The performance measure was the minimum viewing duration required for 75%-correct performance at each location. Thresholds were measured using the step method [8], an adaptive paradigm that uses the subject's response history to adjust a variable parameter across trials and home in on the target performance level. Here, viewing duration was adjusted in increments between a possible minimum of 17 ms (one video frame) and a possible maximum of 250 ms, beyond which the viewer might be able to execute a saccade and fixate the aircraft [9, 10].



Frame 1: Viewer fixates central target

Frame 2: Aircraft is displayed very briefly.

Frame 3: Visual mask is displayed.

Frame 4: Viewer makes 2AFC response.

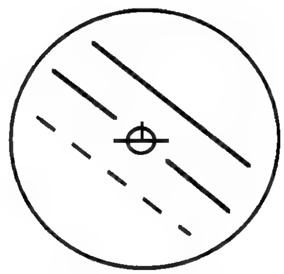
Figure 1

A 32-trial run comprising views of eight aircraft (four fighters and four non-fighters) in four orientations was conducted at each of the twelve screen locations. Trials from the twelve runs were intermixed and presented in random order in each recognition session. The threshold estimates reached at each screen location were recorded for analysis in a 3 (eccentricity) x 4 (quadrant) within-subjects design. Response times were also recorded.

Two performance measures were used to assess individual observers' recognition and manual tracking, respectively. These enabled the subsequent testing in Experiment 2 (in which all subjects would also participate) of whether each individual's near-threshold and manual tracking

ability would be linked to his capacity to perform in a cognitively stressful dual-task situation.

The first performance measure was each subject's mean recognition threshold, averaged across all screen locations. The second measure assessed the viewer's performance on a critical manual tracking task [11, 12]. Viewers used the joystick to control an increasingly-unstable attitude display; the instability that each individual could tolerate before losing attitude control was measured. The display was an "inside-out" roll attitude configuration; a stationary aircraft icon was bracketed by parallel line segments which represented the horizon's orientation (Figure 2). The horizon lines rolled left and right to indicate attitude changes that would result from random stick perturbations. The viewer corrected the attitude by moving the joystick.



"Inside-out" roll attitude indicator, used for critical and subcritical manual tracking tasks.

Figure 2

Performance was assessed using *lambda*, the index of instability. Sessions began with lambda set at 1.5. At this low value, the display was lazy and forgiving of slow correction inputs from the viewer. As the trial progressed, lambda was increased gradually until the viewer lost control, which was defined as a roll angle greater than 180 deg in either direction. After each crash, lambda was recorded and then reset to a lower, manageable value, from which a new trial would begin. Trials continued until 7500 frames (at 15.5 frames/s) had been displayed. The performance measure was each block's median crash lambda value.

In the test session, one recognition and one tracking session were presented. Presentation was counterbalanced across subjects.

Results and Discussion

Aircraft recognition deteriorated with increased eccentricity: Aircraft in the periphery must be displayed for a longer time than targets near the fixation axis, for viewers to classify them successfully, $F(2, 38)=61.3$, $p<.001$ (Figure 3). Furthermore, viewers took longer to effect the classification response for aircraft at the greatest eccentricity value, $F(2, 38)=4.45$, $p<.02$.

Threshold durations for the upper visual field were slightly shorter than lower-visual-field thresholds, and thresholds were shorter for right-field targets than for left-field targets [5, 6]; the quadrant effect was not significant, however. Response latencies were longer in the lower and left visual fields, but this effect was not significant, $F(3, 57)=2.38$, $p=.079$.

Experiment 1 also identified individual performance differences. A central clustering of subject threshold means between 100 and 140 ms was observed. Three subjects fell below this window; these required markedly less viewing time to classify the planes than did the rest of the subjects. Three means were greater than the central cluster, indicating subjects who required substantially more viewing time. It remained to be seen in Experiment 2 whether this measure of the viewer's ability to process briefly presented off-axis targets would predict his ability to perform two tasks at once.

The second individual performance measure, crash lambda, yielded a similar performance distribution for tracking; certain viewers distinguished themselves as better or poorer trackers. Endsley and Bolstad [3] observed that pilots' manual tracking performance was

correlated with performance on a global situational awareness battery, and hypothesized that pilots who possess good tracking skills are able to devote more attention towards situational assessment.

If this principle applied generally to the performance of multiple perceptual tasks, viewers with the best manual tracking ability in a population should suffer the least performance loss when required to switch to dual-task conditions. Experiment 2 tested this also.

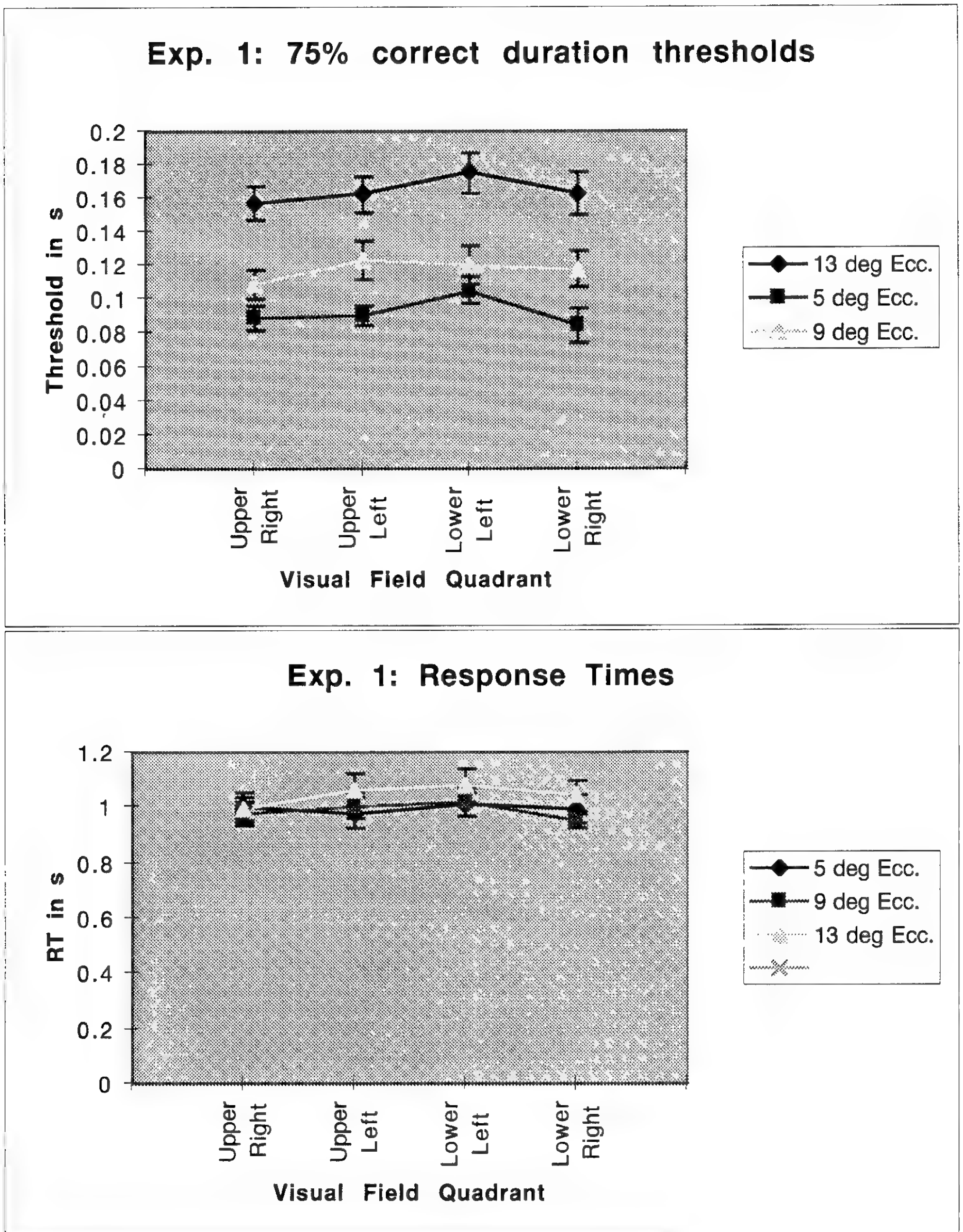


Figure 3

Experiment 2: Does near-threshold recognition (or manual tracking performance) predict dual-task performance?

To test the principle that exceptional ability in processing briefly presented targets confers an advantage for multi-tasking, Experiment 2 tested how performance on the two single tasks would suffer when the subject must perform them concurrently. Objectives were to determine whether viewers' near-threshold and tracking performance would predict their ability to thrive under dual-task conditions, and to determine whether visual field recognition biases would emerge under dual-task workload.

This approach rests on a theoretical assumption that the pilot's capacity to share attention among several tasks is relevant to cockpit performance [2, 3, 4]. Weinstein and Wickens [14] examined this question, and determined that using central and peripheral vision simultaneously to perform two disparate cockpit tasks can cause visual overload.

A reasonable corollary question is whether the viewer's ability to manage this overload distinguishes high- and low-performing pilots. Experiment 2 evaluated dual-task performance as a possible assessment tool. The same twenty subjects practiced performing the tasks from Experiment 1 simultaneously, in three training sessions. They completed single- and dual-task test sessions on the fourth and fifth days. In dual-task conditions, viewers were instructed to treat recognition as primary. The displays and design for the recognition task were as in Experiment 1, except that viewing durations fixed at 83, 100, and 150 ms, respectively, for targets at 5, 9, and 13 deg eccentricities. These durations were the average thresholds at each eccentricity in Experiment 1. The task was constructed thus, to present all viewers with the same objective challenge to recognize near-threshold aircraft as a first priority. This tested the prediction that individuals with the lowest thresholds from Experiment 1 (who could presumably process stimuli with the least investment of attention) would suffer least from the dual-task transition. In contrast, individuals with high thresholds would be taxed more by views of the same duration, so their dual-task performance would be expected to suffer more.

As in Experiment 1, eight aircraft were displayed in four orientations at each screen location. These twelve 32-trial runs, presented randomly, comprised a recognition session. Percent-correct and response latencies were recorded, with number of tasks (single vs dual), eccentricity, and quadrant as within-subject factors. Percent-correct replaced the adaptive thresholds from Experiment 1 because viewing duration was not adjusted across trials. Percent-correct and RT grand means were recorded to measure each subject's performance.

The tracking display also was similar to that in Experiment 1; however, instead of increasing, instability remained constant at a manageable, "subcritical" value, 55% of the subject's peak critical lambda in Experiment 1. Since critical tracking by definition increases instability until the viewer crashes, instructing viewers to treat the task as secondary would induce them simply to abandon that task. Subcritical tracking [12, 15] allowed subjects to complete dual-task sessions without crashing frequently in Experiment 2.

A forcing function imposed disturbances over time to simulate gusts rolling the horizon left and right. Tracking test sessions comprised eight blocks. RMS error for each block was recorded in single- and dual-task conditions. RMS replaced the adaptive lambda measure, because instability remained constant throughout the task. Each subject's dual-task tracking was assessed using the mean proportional increase in RMS error, relative to single task tracking.

In dual-task conditions, the tasks were interdependent. Negative feedback was included to warn subjects who were failing to respond, or were responding incorrectly, that they should protect the primary task. Conversely, in order to complete the recognition task, subjects must not crash.

Correlations were run on each subject's individual performance measures, to determine whether predictive links would emerge between a subjects' single-task near-threshold recognition and tracking, and his dual-task performance (Table 1).

Expl: Threshold	Mean 75%-correct duration threshold as measured in Experiment 1, across eccentricity and quadrant levels.
Expl: Lambda	Median peak lambda achieved in Experiment 1.
SingleCorrect	Correct aircraft classifications (percent), single task condition, for all screen locations.
DualCorrect	Correct aircraft classifications (percent), dual-task condition, for all screen locations.
CorrectDelta	(DualCorr - SingleCorr)
SingleRT	Mean classification response time, single task condition, all locations.
DualRT	Mean classification response time, dual-task condition. all locations.
RTDelta	(DualRT - SingleRT)
SingleRMS	RMS roll tracking error, single task condition
DualRMS	RMS roll tracking error, dual-task condition
RMSDelta	(DualRMS - SingleRMS) / SingleRMS

Table 1. Individual Performance Measures

Results and Discussion

Effects of task, eccentricity, and quadrant

There were more correct responses in single- than dual-task conditions, F(1, 19)=51.4, p<.05 (Figure 4). There was no eccentricity effect on percent-correct, which indicates that the thresholds from Exp. 1 (on which our display durations were based) and percent-correct were mutually consistent; thresholds registered the greater viewing time required at wider eccentricities. There was no effect of quadrant on percent-correct. No significant task x eccentricity, task x quadrant, quadrant x eccentricity, or three-way interaction was found.

Response latencies were *shorter* in dual-task conditions, F(1, 19)=13.5, p<.005, and increased with eccentricity, F(2, 38)=18.4, p<.001. Quadrant influenced RT, with viewers responding sooner to aircraft in the upper and right visual fields, F(3, 57)=4.83, p<.01. Eccentricity x quadrant interaction was apparent but missed significance, F(6, 114)=2.17, p=.051. No task x eccentricity, task x quadrant, or three-way interaction was found for RT.

RMS error was greater in dual-task conditions, F(1, 19)=76.9, p<.001 and varied with presentation block, F(7, 133)=5.75, p<.001. There was an interaction between number of tasks and block, F(7, 133)=2.73, p<.05.

While adding a second task was apparently less harmful for recognition than for tracking, the fact that recognition was degraded indicates that subjects were not totally successful in preserving the primary task. The dual-task effect on accuracy showed that visual overload hindered recognition even when viewers were instructed to give it priority.

Spatial biases have been observed in visual attention, which suggests that resource limitations affect performance differentially throughout the visual field. Previc and Blume [6] proposed a visual performance contour for a distance-biased attentional system that favors the upper right quadrant, and whose evolutionary function is to search for and recognize objects. The main quadrant effect on response times was consistent with such a spatial attention bias: Viewers took longer to classify aircraft in the upper and right visual fields. Furthermore, a non-significant interactive trend (F(3, 57)=5.79, p=.086) suggested that recognition was hindered in a spatially biased way in high-workload conditions.

The main effect of the number of tasks on RT is paradoxical, and offers a clear demonstration

of the complexities inherent in using RT for performance assessment. The importance of classifying aircraft correctly (as opposed to quickly) was stressed in the instructions, as was the priority of this task. Nevertheless, an apparent speed-accuracy trade-off occurred whereby viewers responded slightly less accurately, but faster, than in single-task conditions.

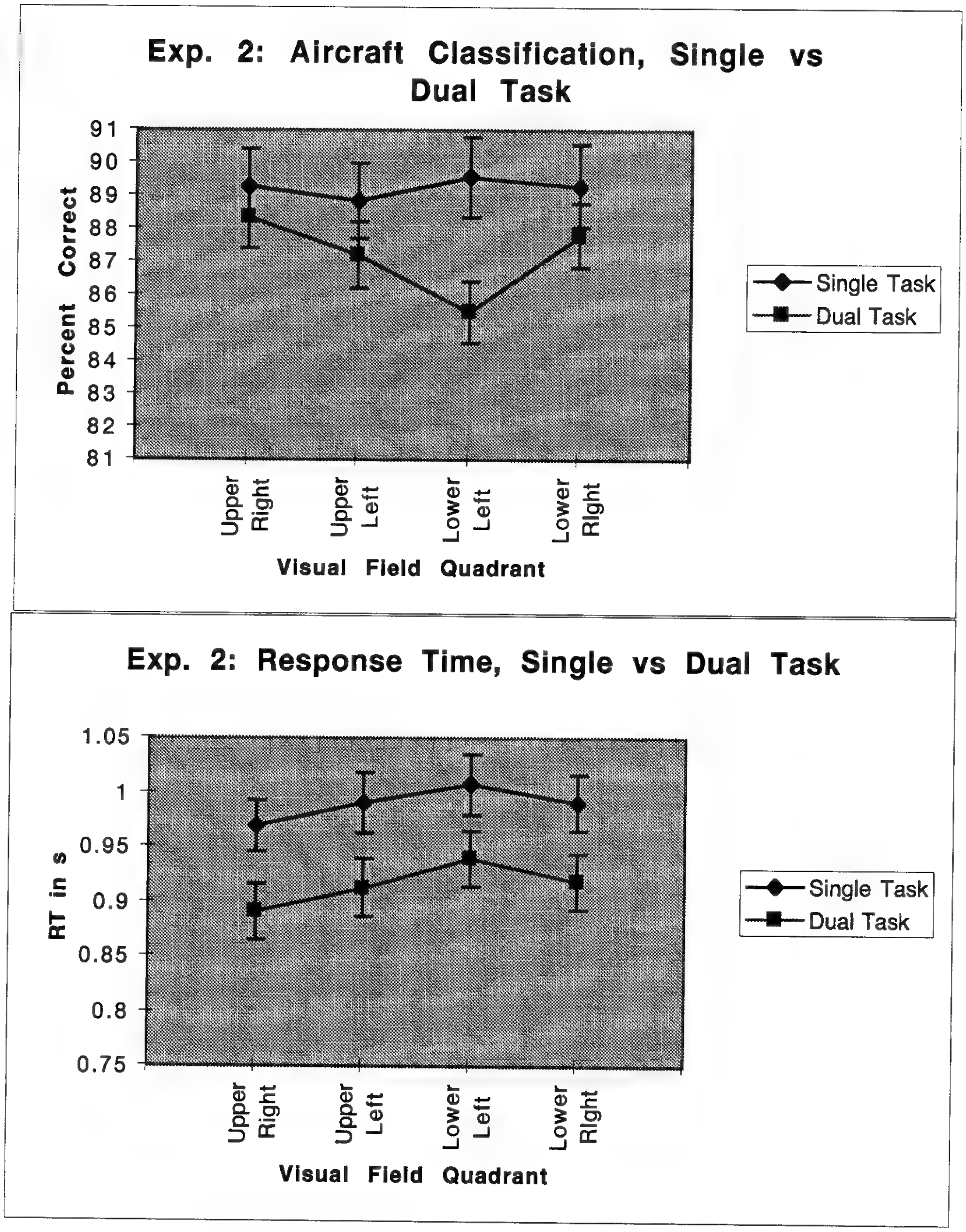


Figure 4

Manual attitude control, which viewers were instructed to let slip if necessary, suffered considerably under dual-task loading and deteriorated in the later test blocks. Like the recognition data, the tracking data are consistent with an attention model in which a visual resource

pool is shared across disparate tasks performed in central and peripheral vision [14]. The relatively greater degradation in tracking indicates that viewers enjoyed some success at managing visual overload to preserve the task with greater utility.

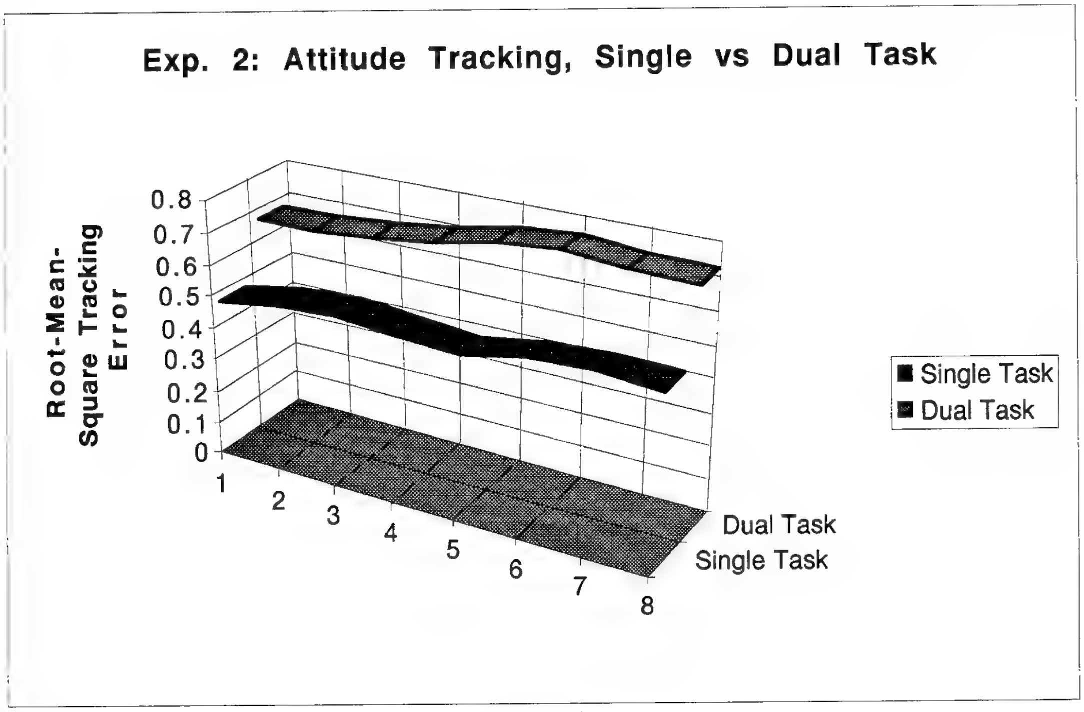


Figure 5

Dual-task performance in Experiment 2 can be described in the following manner. The viewer performs the default tracking task continuously. Tracking is characterized by lag and gain characteristics, which describe the viewer's ability to counter perturbations within a frequency envelope [12, 16 (pp. 486-488)]. Maintaining attitude consumes resources from a visual processing pool. At intervals, aircraft targets require the viewer to switch resources to the primary task. This can be characterized as the opening of an attention gate [17, 18] which admits recognition information at the relevant peripheral location. Eccentricity and visual field effects result from variations in the time course of the opening; for example, if visual information is degraded from targets at large eccentricities, or in the lower visual field.

Since viewers were instructed to expend effort on tracking only after responding confidently to the aircraft, good performance on both tasks necessitated efficient attention switching. The lack of resources for the secondary task would effectively relax tracking criteria whenever an aircraft was displayed. This allowed attitude to topple around more, and recover only after the viewer diverted resources back to tracking. Ragged attitude control offered an effective incentive to switch quickly back to tracking at the appropriate time; consequently, dual-task response times were shorter.

Individual performance measures, single and dual

We propose that the efficiency of the viewer's attention switching strategy determined the viewer's dual-task performance. The array of individual performance measures, and the correlations between them, can be interpreted to support this claim (Table 2).

The first obvious finding was that thresholds from Experiment 1 correlated significantly with percent-correct in dual-task conditions. Certain viewers needed less viewing time than others to classify aircraft reliably. These viewers identified more aircraft correctly at fixed

durations, an advantage that survived the transition to dual-task conditions.

Measure 1	Measure 2	Pearson r	Spearman ρ
SingleCorrect	Expl: Threshold	**-.619***	**-.728***
DualCorrect	Expl: Threshold	**-.628***	**-.609***
CorrectDelta	Expl: Threshold	-.298	-.097
SingleCorrect	Expl: Lambda	.223	
DualCorrect	Expl: Lambda	.433	
CorrectDelta	Expl: Lambda	**.492***	
RMSDelta	Expl: Threshold	.308	.128
RMSDelta	Expl: Lambda	-.161	
RMSDelta	SingleCorrect	-.168	
RMSDelta	SingleRT	.349	.204
RMSDelta	CorrectDelta	-.340	
DualCorrect	SingleCorrect	**.848***	
RTDelta	Expl: Threshold	**-.579***	**-.695***
RTDelta	Expl: Lambda	.178	
RTDelta	SingleCorrect	**.600***	
RTDelta	DualCorrect	**.460***	
RTDelta	DualRMS	.092	
DualCorrect	DualRT	-.134	

Table 2. Exp. 2: Correlations Between Performance Measures

However, the correlation was nonsignificant between viewers' recognition thresholds, as measured in Experiment 1, and the attitude control measure, which was the proportional increase in RMS tracking error from single to dual-task tracking. Overall, the subcritical tracking data offered little evidence to indicate that an individual's near-threshold processing predicts his dual-task performance.

In contrast to subcritical tracking, critical tracking proved to be a surprisingly interesting measurement tool. The experiment's second goal was to test whether viewers who are good at manual tracking are better equipped to handle stressful multiple-task conditions. Critical tracking, which pressures the viewer continuously until he crashes, appears to involve the same attention resources that underlie

reliable near-threshold aircraft classification in dual-task conditions: Viewers who could withstand greater tracking instability, as measured in Experiment 1, were better at preserving dual-task recognition. This finding is consistent with the hypothesis that good tracking is a characteristic of pilots who can use attention efficiently, to establish and update a global situational percept [3].

An additional unexpected finding was the group of correlations that linked the change in response time between single- and dual-task conditions and the three indices of recognition performance. Surprisingly, viewers who were good at aircraft classification, as indicated by duration thresholds ($r=-.58$), single-task percent-correct ($r=.60$), and dual-task percent-correct ($r=.46$) were less likely to speed their responses on that task in high-workload conditions. A possible explanation is that viewers' awareness of the effort required to switch attention efficiently might have influenced their allocation strategy and performance. High-threshold viewers (who were less proficient at recognition) may have stolen more time from the recognition task because they were more acutely aware of the resources they were spending on it, and more aware that this expenditure taxed their ability to maintain the tracking task. In the classroom analogy of attention switching [18, 19], this awareness constitutes a cue to dash from aircraft class early ("where I'm having a hard time anyway"), to attend manual tracking ("which I'm quite sure I can pass"). This notion of situational arousal, defined as an awareness of multitask workload, might prove to be a useful construct for predicting a viewer's SA; in the present dual-task situation, it appears that the extent of such arousal would be an SA liability, not a benefit.

CONCLUSIONS

Experiments 1 and 2 identified three aspects of near-threshold and manual tracking performance that are relevant to pilot performance in stressful cockpit situations.

Adaptive threshold estimation [8] is useful for assessing viewers' recognition of real-world targets. Measured thresholds were found to predict viewers' ability to recognize off-axis aircraft in dual-task conditions. Furthermore, viewers who processed near-threshold targets most efficiently were less likely to abandon that task early under stressful dual-task conditions. However, no strong evidence was found to indicate that low visual recognition thresholds are associated with a general multi-tasking competence that supports the concurrent performance of a second, disparate cockpit task.

The second class of findings comprised a surprising group of correlations that were observed between critical tracking and dual-task recognition. At present, it is unclear whether viewers who were superior at critical tracking preserved dual-task recognition better because of superior switching strategies, or some other ability. However, it appears that the critical tracking paradigm, which pressures the viewer inexorably until he loses control, tapped a competence that also supported the maintenance of visual attention in the periphery. Critical tracking is a promising tool for further examination of the competence framework that underlies multitasking in the cockpit. An objective for future research is to test whether critical tracking is related to other multitasking performance measures, including global SA assessment tools [3, 20,21].

The third class of findings suggested that the worst place for a fleeting target to show up is in the pilot's lower left visual field, especially if he or she is busy. Whereas a proposed model of a spatially biased extrapersonal attention system had been tested using visual search, this class of findings extended the model to the recognition of fleeting aircraft. Importantly, targets were small and were focussed at infinity, as would be the case with real aircraft in the distance periphery.

In sum, the potential relevance of these findings for cockpit performance is manifold. Further research is needed, to refine and clarify the applicability of these measures for assessing global situation perception.

References

1. Hartman, B. O., & Secrist, G. E., "Situational awareness is more than exceptional vision," Aviation, Space, and Environmental Medicine, 62, 1991, pp. 1084-9.

2. Damos, D. L., "Residual attention as a predictor of pilot performance," Human Factors, 20, 1978, pp. 435-440.

3. Endsley, M. R., & Bolstad, C. A., "Individual differences in pilot situational awareness," The International Journal of Aviation Psychology, 4, 1994, pp. 241-264.

4. North, R. A., & Gopher, D., "Measures of attention as predictors of flight performance," Human Factors, 18, 1976, pp. 1-14.

5. Previc, F. H., "Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications," Behavioral and Brain Sciences, 13, 1990, pp. 519-575.

6. Previc, F. H., & Blume, J. L., "Visual search asymmetries in three-dimensional space," Vision Research, 33, 1993, pp. 2697-2704.

7. Hochberg, J., "Perception, 2nd ed.," Englewood Cliffs, NJ: Prentice-Hall, 1978.

8. Simpson, W. A., "The step method: a new adaptive psychophysical procedure," Perception & Psychophysics, 45, 1989, pp. 572-576.

9. Hallett, P. E., "Eye movements," In K. Boff, J. Thomas, & L. Kaufman (Eds.), "Handbook of perception and human performance, Vol. 1," New York: Wiley, 1986, pp. 10.1-10.112.

10. Heywood, S., & Churcher, J., "Structure of the visual array and saccadic latency: Implications for oculomotor control," Quarterly Journal of Experimental Psychology, 32, 1980, pp. 335-341.

11. Jex, H. R., McDonnell, J. D., & Phatak, A. V., "A 'critical' tracking task for manual control research," IEEE Transactions on Human Factors in Electronics, HFE-7, 1966, pp. 138-145.

12. Kenyon, R. V., & Kneller, E. W., "The effects of field of view size on the control of roll motion," IEEE Transactions on Systems, Man, and Cybernetics, 23, 1993, pp. 183-193.

13. Secrist, G. E., & Hartman, B. O., "Situational awareness: the trainability of the near-threshold information acquisition dimension," Aviation, Space, and Environmental Medicine, 64, 1992, pp. 885-892.

14. Weinstein, L. F., & Wickens, C. D., "Use of nontraditional flight displays for the reduction of central visual overload in the cockpit," The International Journal of Aviation Psychology, 2, 1992, pp. 121-142.

15. Previc, F. H., Kenyon, R. V., Boer, E. R., & Johnson, B. H., "The effects of background visual roll stimulation on postural and manual control and self-motion perception," Perception & Psychophysics, 54, 1993, pp. 93-107.

16. Wickens, C. D., "Engineering psychology and human performance, 2nd ed.," New York: Harper Collins, 1992.

17. Reeves, A., & Sperling, G., "Attentional gating in short-term visual memory," Psychological Review, 93, 1986, pp. 180-206.

18. Sperling, G., & Dosher, B. A., "Strategy and optimization in human information processing," In K. Boff, J. Thomas, & L. Kaufman (Eds.), "Handbook of perception and human performance, Vol. 1," New York: Wiley, 1986, pp. 2.1-2.65.

19. Sperling, G., & Melchner, M. J., "Visual seach, visual attention, and the attention operating characteristic," In J. Requin (Ed.), "Attention and performance VII," Hillsdale, N. J.: Erlbaum, 1978.

20. Endsley, M. R., "A methodology for the objective measurement of situational awareness," In "Situational Awareness in Aerospace Operations," Copenhagen, Denmark: North Atlantic Treaty Organization, AGARD, 1989, pp. 1.1-1.9.

21. Sarter, N. B., & Woods, D. D., "Situation awareness: a critical but ill-defined phenomenon," The International Journal of Aviation Psychology, 1, 1991, pp. 45-57.

DEVELOPMENT OF TECHNIQUES TO IDENTIFY INDIVIDUALS WITH SUPERIOR POTENTIAL FOR SITUATIONAL AWARENESS

R.E. King
C.F. Flynn
Neuropsychiatry Branch
Clinical Sciences Division
Aerospace Medicine Directorate
Armstrong Laboratory (AFMC)
2507 Kennedy Circle
Brooks AFB, Texas 78235-5117
United States of America

SUMMARY

Certain cognitive abilities and personality traits may be conducive to the development of situational awareness. The United States Air Force Neuropsychiatrically Enhanced Flight Screening (N-EFS) program screens pilot candidates before they enter undergraduate pilot training (UPT). The Multidimensional Aptitude Battery (MAB), a highly timed IQ test, and CogScreen, a computer-based cognitive test, are now administered to all UPT candidates. CogScreen measures attention, memory, visual scanning, response speed, visual-spatial orientation, capacity for divided attention, and resistance to response interference. CogScreen approximates and measures response to the multiple, competing activities requiring prompt and prioritized attention. Traditional neuropsychological tests do not gauge the subtle abilities that performance in a high-demand environment requires on account of their clinical, rather than occupational, emphasis. The Personal Characteristics Inventory (PCI) and the Revised NEO-Personality Inventory (NEO-PI-R) comprise the optional (requiring participant informed consent) portion of N-EFS. PCI measures decision-making strategies and interpersonal style, gauging potential for effective crew resource management. The NEO-PI-R may suggest optimal personality styles for developing situational awareness in particular aircraft. Preliminary results suggest that prospective pilots have a wide range of intelligence and cognitive ability (from average to very superior) and distinct personality styles. Testing results captured prior to the commencement of training will be compared to occupational outcome (whether or not the candidate became a mission-ready pilot) to assess their predictive value in the development of situational awareness.

TEXT

Selecting individuals with the best and fastest cognitive abilities and personalities suited for effective crew resource management may be an effective method to better ensure aviators develop the degree of situational awareness (SA) needed for particular airframes and missions. Formerly, pilots' baseline cognitive ability was not assessed by the U.S. Air Force (USAF, 13). Quick and accurate information processing skills and good judgment, however, are of paramount importance to the modern aviator. Hartman and Secrist (16) assert that SA is more than exceptional vision; it is mostly cognitive. They identify other important attributes, including strong motivation to fly, competitive aggressiveness, and ability to tolerate and manage high levels of stress. Hartman and Secrist challenge the aeromedical community to identify the abilities that lead to superior SA. They assert: "Selection of individuals that possess abilities beneficial to SA certainly makes sense if it can be accomplished" (16, p. 24).

The USAF Surgeon General has answered this challenge by supporting the Neuropsychiatrically Enhanced Flight Screening (N-EFS) program. N-EFS assesses pilot candidates before they enter undergraduate pilot training (UPT), during their flight screening. N-EFS is investigating, with consenting student pilots, if it's possible to use psychological testing to *prospectively* and *validly* predict who will become mission ready. The criterion of whether or not an aviator becomes mission ready is directly related to whether or not that individual is able to develop adequate SA for their given airframe.

Successful pilots must be cognitively and psychologically "fit" to preserve SA while executing complex job demands in an unforgiving environment. The testing data collected by N-EFS will establish a range of cognitive attributes of pilots and aspiring pilots, then later compare it to occupational outcome (whether or not the candidate becomes mission ready, i.e., a fully qualified left-seat pilot) to help understand the qualities that lead to superior SA. The former USAF Chief of Staff, General Merrill A. McPeak, asked if the ability to develop SA is correlated with gender. N-EFS stores data with gender as a variable (necessitated by the published personality norms that

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

are based on gender, 7).

The contribution of personality and judgment to pilot success is poorly defined, while sensory-motor skills have been more frequently scrutinized (3, 4, 25). Further, previous selection studies (4, 28, 29, 31) used the criterion of success as completion of undergraduate pilot training (UPT), rather than whether or not the candidate achieved actual mission-ready status. Using this short-term criterion plagues research results with the "honey-moon effect;" pilot students attempt to look their best and sustain a high level of performance, but only in the short-run (17). Measures that focus on who *can* (aptitude) finish pilot training, misses information on who *will* (motivation) finish pilot training and evolve into effective military resources. Beyond student pilot data, mission-ready military pilot psychometric norms are severely limited or based on very small, specialized populations (2, 10, 12, 27, 36).

Demands for greater aircraft speed challenged airframes; now speed of information processing is becoming a critical skill for the new generation of military pilot. As military forces shrink while mission demands expand and diversify, selecting pilots for stress resilience and speed of information processing is paramount. The successful aviator must choose the most critical data from a myriad of cockpit instruments to maintain flight safety and achieve mission completion (16). How aviators optimally process this information is now becoming an important operation to define. This "extraordinary awareness of the total flight environment" is but one definition of SA (33).

Exactly what SA is, how one is to measure it, and whether it is an inherent trait or a trainable skill are open questions. Some common components in varying definitions of SA are: the capability to compose a multitude of data bits into a composite understanding (8), relating this understanding to aircraft and environment (20), anticipating future actions by matching information to known patterns (16), and consequently prioritizing one's actions to maximize inherent advantages. Hartman and Secrist (16) note that no conclusive evidence exists whether it is possible to select personnel who are especially prone to developing effective skilled memory but note the need for tests for examining the development of skilled memory (16, p. 22).

Vidulich and coauthors (32) suggest SA is a combination of perceptual motor and cognitive skills. SA requires the capability to sort through layers of information in a time-critical period and then effectively prioritizing task completion. In addition, they note attentional capabilities, personality reactions to stress, and emotional control have a critical impact on SA. Personality traits and cognitive capabilities can be studied through the N-EFS program. Comparing results of successful pilots (those who successfully upgrade to full mission-ready status) to unsuccessful (those who fail to upgrade) candidates could identify areas important to SA.

Glimpses of the USAF aviator population have suggested that they tend to have superior intelligence as measured by IQ tests (10, 12). Aviators with good SA, however, may have other cognitive gifts, such as superior dual-tasking ability or rapid information processing abilities (8, 32). The Multidimensional Aptitude Battery (MAB; 21, 22), a highly timed IQ test, and CogScreen, a computer-based cognitive test, are now administered to all UPT candidates. Although the MAB and CogScreen may both be measuring "g" (general intelligence), they share, at most, less than 9% of their variance (G. Kay, Ph.D., unpublished manuscript, 1994). Administering CogScreen in addition to the MAB may add incremental validity in determining who is likely to become a mission-ready aviator.

The MAB is a multiple choice IQ test that Armstrong Laboratory helped computerize. The MAB is administered in ten seven-minute blocks and yields the following scores:

<u>Verbal IQ</u> (Crystallized ability - results from interaction with the culture)

Subtests:
Information (Fund of knowledge, long-term memory)

Comprehension (Ability to evaluate social behavior)

Arithmetic (Reasoning and problem-solving ability)

Similarities (Flexibility, adjustment to novelty, abstract thought, long-term memory)

Vocabulary (Openness to new information, capacity to store, categorize, and retrieve words and verbal concepts previously learned)

<u>Performance IQ</u> ("Fluid" ability - independent of education and experience, capacity for learning and problem solving)

Subtests:
Digit Symbol (Adaptation to a new set of demands, learning coding, performing visual-motor activity)

Picture Completion (Identifying important missing elements in a picture, knowledge of common objects)

Spatial (Ability to visualize abstract objects in different positions)

Picture Completion (Ability to identify a meaningful

sequence, social intelligence and insight in others' behavior)

Object Assembly (Visualization skills and perceptual analytical skills needed to identify a meaningful object from a left-to-right sequence)

Full Scale IQ (General aptitude - comprised of all subtests)

CogScreen is a computer-based, self-administered neuropsychological screening tool, which requires a light-pen with interface software, and a carathoid ray tube (CRT) monitor. CogScreen is also being used with commercial aviator populations to assess learning and memory retention and these abilities are being correlated with cockpit performance (18, 19). In comparison to traditional neuropsychological assessment, which is administrator labor-intensive and time-consuming, CogScreen may ultimately prove superior in identifying cognitive subtleties that are key to SA in a high-demand environment. CogScreen measures attention, memory, visual scanning, response speed, visual-spatial orientation, capacity for divided attention, and resistance to response interference. CogScreen is heavily memory dependent and is able to very accurately time a subject's speed of response. CogScreen approximates and measures response to the multiple, competing activities requiring prompt and prioritized attention. Traditional neuropsychological tests do not gauge the subtle abilities that performance in a high-demand environment requires due to their clinical, rather than occupational, emphasis. CogScreen can be administered in a group setting by using headphones and disabling the external speaker of the computer (thereby delivering auditory feedback to a participant without confusing other participants). The computer microprocessor has placed added demands on the high performance aviator; it may also help us identify the candidates with the best potential for SA.

CogScreen:

Backward Digit Span (Attention, working memory, verbo-sequential processing)

Math Problems (Working and long-term memory, logical reasoning)

Visual Sequence Comparison (Attention, working memory, and verbo-sequential processing)

Symbol Digit Coding (Attention, visual scanning, working memory, verbo-sequential processing)

Symbol Digit Coding - Immediate Recall (Immediate recall)

Matching to Sample (Visuo-spatial memory, response speed)

Manikin Figures (Visuo-spatial orientation, ability to rotate mental images, long-term memory)

Divided Attention Test (Speed and accuracy of responding. In dual task mode: Divided attention, working memory, visual-spatial processing, verbo-sequential processing)

Auditory Sequence Comparison (Attention, working memory, verbo-sequential processing)

Pathfinder (Verbo-sequential processing, working memory, attention, ability to systematically apply rules)

Symbol Digit Coding - Delayed Recall (Memory and recall)

Shifting Attention Test (Concept formation, conceptual flexibility, deductive reasoning, response interference

CogScreen may be particularly well suited for identifying which candidates will go on to develop exceptional SA in flying due to the need Hartman and Secrist (16) identify for pattern recognition, or the ability to quickly size up a situation and accurately process it. While fluid intelligence and memory are intuitively obvious prerequisites for SA, verbal (crystallized) intelligence may also be vital. Hartman and Secrist champion the ability to use near-threshold information, although they are unsure whether or not it is an ability that should be "measured in the course of selection, or is a general skill that can be trained in the student pilot population, (or if) it is some combination of the two" (16, p. 22). They continue: "whatever spatial model the pilot might be maintaining to perform the task will have to be supplemented by verbal information from other pilots or controllers, so testing verbal abilities and capacities also makes sense" (16, p. 24). Hartman and Secrist also cite the importance of stress resistance, due to the high information loads and physical stressors of modern flight. A strong reaction to stress and lack of emotional control disrupt SA. In any event, regardless of any inherent capacity (trait quality) of a pilot for SA, the *quality* (state) of SA may change from one flight to the next, due to sleep deprivation, psychoactive agents, lack of specific or recent training, or transient emotional states.

Vidulich and coauthors (32) emphasize the importance of personality, as above average self-discipline leads to healthy questioning, a larger fund of aviation knowledge, a more professional attitude towards flying, vigilance to radios, maintenance, and weather, and a pilot who is always thinking ahead of his/her aircraft. "Staying ahead of the airplane," thinking of the next

contingency, is clearly key to SA.

N-EFS, therefore, solicits data on personality and judgment as measured by the Personal Characteristics Inventory (PCI, 15) and the Revised NEO-Personality Inventory (NEO-PI-R, 7), the research (optional) portion of the N-EFS battery. PCI measures decision-making strategies and interpersonal style, gauging potential for effective crew resource management (CRM, 15). Vidulich and coauthors (32) include leadership, communicative ability, and interpersonal skills as the ingredients of CRM and assert good SA flows from good CRM. The PCI consists of 254 questions presented in a Likert format ("Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree"). Aircrew are categorized into eight groups ranging from the "right stuff" to the "wrong stuff" in cockpit crew coordination. Already widely used in civilian aerospace operations, groups generated by the PCI can now be correlated more specifically to SA.

The NEO-PI-R gauges normal personality functioning based on the five-factor model of normal personality (6, 7, 14). Consisting of 240 questions, also in a Likert scale format, it is not being used to identify psychopathology (a *select-out* function) in N-EFS. Aviators as a group appear to harbor very little personality pathology (1, 24, 30). Rather, the NEO-PI-R may identify whether some personality traits are more predictive of success in military aviation, which would aid future pilot selection (*select-in*). The NEO-PI-R may suggest optimal personality styles for pilots in particular aircraft. For example, excessive "excitement-seeking" tendencies may be contraindicated in pilots flying aircraft with high cognitive workload requirements. Pilots with stimulus filters may be more appropriate for these demanding aircraft. The link between performance and optimal arousal has long been known (35).

<u>Revised NEO Personality Inventory (NEO PI-R):</u>
Domains (numeric)
    Facets (alphanumeric)

Neuroticism (N)
- *Anxiety (N1)*
- *Angry Hostility (N2)*
- *Depression (N3)*
- *Self-Consciousness (N4)*
- *Impulsiveness (N5)*
- *Vulnerability (N6)*

Extraversion (E)
- *Warmth (E1)*
- *Gregariousness (E2)*
- *Assertiveness (E3)*
- *Activity (E4)*
- *Excitement-Seeking (E5)*
- *Positive Emotions (E6)*

Openness (O)
- *Fantasy (O1)*
- *Aesthetics (O2)*
- *Feelings (O3)*
- *Actions (O4)*
- *Ideas (O5)*
- *Values (O6)*

Agreeableness (A)
- *Trust (A1)*
- *Straightforwardness (A2)*
- *Altruism (A3)*
- *Compliance (A4)*
- *Modesty (A5)*
- *Tender-Mindedness (A6)*

Conscientiousness (C)
- *Competence (C1)*
- *Order (C2)*
- *Dutifulness (C3)*
- *Achievement Striving (C4)*
- *Self-Discipline (C5)*
- *Deliberation (C6)*

(Facets are components of their respective domains. Interested readers are referred to reference 7 for a complete description of domain and facet characteristics.)

Candidates for undergraduate pilot training present themselves with a wide variety of baseline intellectual ability and personality styles (23). Will these vast differences lead to differential ability in developing SA and thus becoming mission ready?

The smaller air forces of the future will rely increasingly on the accuracy and skill of fewer aviators flying fewer, but more sophisticated and unforgiving, aircraft. Defining the psychological and neuropsychological characteristics of successful aircrew could yield greater understanding of the elements of SA. This research has not been accomplished to date in the USAF, perhaps because it is difficult to obtain psychometric information from aircrew as they may fear test results will lead to medical disqualification.

Aviator occupational norms for psychological testing need to be defined, as they currently do not exist. Norms could help identify psychological characteristics of those candidates who become mission-qualified pilots by displaying good SA. Data from the MAB, CogScreen, PCI, and NEO-PI-R will shed more light on the cognitive and emotional aspects of good SA.

References

1. Adams RR, Jones DR. Healthy motivation to fly: no psychiatric diagnosis. Aviat. Space Environ. Med. 1987; 58:350-4.

2. Ashman A, Telfer R. Personality profiles of pilots. Aviat. Space Environ. Med. 1983; 54:940-3.

3. Carretta TR. Basic attributes tests (BAT) system: development of an automated test battery for pilot selection. USAF Human Resources Laboratory TR-87-9, 1987.

4. Carretta TR. USAF pilot selection and classification systems. Aviat. Space Environ. Med. 1989; 60:46-9.

5. Chidester TR, Helmreich RL, Gregorich SE, Geis CE. Pilot personality and crew coordination: Implications for training and selection. Int. J. Aviat. Psychology 1991; 1:25-44.

6. Conoley JC, Kramer JJ. The tenth mental measurements yearbook. Lincoln, Nebraska: The University of Nebraska Press, 1989.

7. Costa PT, McCrae RR. Professional manual: revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources, Inc., 1992.

8. Damos D. The effects of high information processing loads on human performance. In: Proceedings of the 7th behavioral technology conference and exposition, 1988 Oct. Warrendale, PA: SAE, Inc. 1989; 51-4.

9. Endsley MR, Belstad CA. Human capabilities and limitations in situation awareness. IN: AGARD, Combat Automation for Airborne Weapons Systems: Man/Machine Interface Trends and Technology. Neuilly-Sur-Seine, France: NATO-AGARD; N-93-28850, 1993.

10. Fine M, Hartman BO. Psychiatric strengths and weaknesses of typical air force pilots. USAF School of Aerospace Medicine TR-68-121, 1968.

11. Flynn CF, Sipes WE, Grosenbach MJ, Ellsworth J. Field test of a computer-driven tool to measure psychological characteristics of aircrew. USAF Armstrong Laboratory TR-92-171, 1992.

12. Flynn CF, Sipes WE, Grosenbach MJ, Ellsworth J. Top performer survey: computerized psychological assessment in aircrew. Aviat. Space Environ. Med. 1994; 65(5, Suppl.):A39-44.

13. Gnan M, Flynn CF, King RE. Psychological pilot selection in the U.S. Air Force, the Luftwaffe, and the German Aerospace Research Establishment. Armstrong Laboratory Aerospace Medicine Directorate Technical Report 1995-0003. U.S. Government Printing Office, Washington, D.C.

14. Goldberg LR. The development of markers for the big-five factor structure. Psychological Assess. 1992; 4:26-42.

15. Gregorich S, Helmreich RL, Wilhelm JA. Personality based clusters as predictors of aviator attitudes and performance. In: RS Jensen (ed). Proceedings of the 5th International Symposium on Aviation Psychology, (1989). Vol II. Columbus, OH, 1989: 686-91.

16. Hartman BO, Secrist GE. Situational awareness is more than exceptional vision. Aviat. Space Environ. Med. 1991; 62:1084-9.

17. Helmreich RL, Sawin LL, Carsrud AL. The honeymoon effect in job performance: Delayed predictive power of achievement motivation. J. Applied Psychology 1986; 71:1085-8.

18. Hordinsky J. Testing for brain damage. Aviation Safety Journal. 1992; (2):18-20.

19. Horst RL, Kay GG. Personal computer-based tests of cognitive function for occupational medical certification. In: Proceedings of the 6th International Symposium on Aviation Psychology, 1991 Apr 29-May 2. Vol 2. Columbus OH: Ohio State University, 1991: 734-9.

20. Houck MR, Whitaker LA, Kendall RR. An information processing classification of beyond-visual-range air intercepts. USAF Armstrong Laboratory TR-93-61, 1993.

21. Jackson DN. Multidimensioal Aptitude Battery manual. Ontario, Canada: Research Psychologists Press, Inc, 1984.

22. Jackson III DN, Barton CF, Blokker HC. User's manual for the Multidimensional Aptitude Battery (MAB) software program. Ontario, Canada: Research Psychologists Press, Inc, 1992.

23. King RE, Flynn CF. Continuing the quest to define and measure the "right stuff": Neuropsychaitrically Enhnanced Flight Screening (N-EFS). Aviat. Space Environ. Med. In Press.

24. King RE. Assessing aviators for personality pathology with the Millon Multiaxial Clinical Inventory (MCMI). Aviat. Space Environ. Med. 1994; 65:227-31.

25. Long GE, Varney NC. Automated pilot aptitude measurement system. USAF Human Resources Laboratory TR-75-58, 1975.

26. Pedersen LA, Allan KE, Lave FJ, Johnson JR, Siem FR. Personality theory for aircrew selection and classification. USAF Armstrong Laboratory TR-92-21, 1992.

27. Picano JJ. Personality types among experienced military pilots. Aviat. Space Environ. Med. 1991; 62:517-20.

28. Retzlaff PD, Gibertini M. Objective psychological testing of U.S. Air Force officers in pilot training. Aviat. Space Environ. Med. 1988; 59:661-63.

29. Siem FM. Predictive validity of an automated personality inventory for Air Force pilot selection. Int. J. Aviat. Psychology 1992; 2 (4):261-70.

30. Sipes W, Moore J, Caldwell L. The MMPI: A look for military pilot norms. In: Proceedings of the 33rd conference of the Military Testing Association, 1991 Oct; 1991:429-34.

31. Street DR, Helton KT, & Nontasak T. An evaluation of personality testing and the five-factor model in the selection of landing craft air cushion vehicle crew members. Nav. Aero. Med. Res. Lab. TR-93-1385, 1993.

32. Vidulich M, Dominguez C, Vogel E, McMillan G. Situational awareness: Papers and annotated bibliography AL/CF-TR-1994-0085.

33. Vidulich MA, Stratton M, Crabtree M, Wilson G. Performance-based and physiological measures of situational awareness. Aviat. Space Environ. Med. 1994; 65(5, Suppl):A7-A12.

34. Yacavone DW. Mishap trends and cause factors in naval aviation: A review of naval safety center data, 1986-90. Aviat. Space Environ. Med. 1993; 64:392-5.

35. Yerkes RM Dodson JD The relationship of strength of stimulus to rapidity of habit-formation. Jornal of Comparitive Neurology. 1908; 18:459-482.

36. Youngling EW, Levine SH, Mocharnuk JB, Weston LM. Feasibility study to predict combat effectiveness for selected military roles: fighter pilot effectiveness. St Louis, MO: McDonnell Douglas Astronautics Company-East, MDC-E1634, 1977.

Situational Awareness, Trust, and Compatibility: Using Cognitive Mapping Techniques to Investigate the Relationships between Important Cognitive Systems Variables

Taylor, R.M., Shadrake, R. Haugh, J., and Bunting, A.
DRA Centre for Human Sciences
CH3, F131 Building — Room 43
Farnborough, Hants GU14 6SZ
United Kingdom

1. SUMMARY

Situational awareness (SA), trust, and compatibility are considered as variables associated with cognitive systems. Three studies are reported investigating the relationships between these variables using experimental methods, subjective ratings, and cognitive mapping techniques. In the first study, a computer simulation of an Air Traffic Control (ATC) task was used to investigate the relationship between task performance and subjective estimates of situational awareness using the Situational Awareness Rating Technique (SART). The results show a strong association between rated SA and performance, and provide evidence of the predictive power of a unitary SART index. In the second study, a simulated aircraft task environment was used to investigate the effects of unreliable computer aiding on task performance, and ratings of SA, and of attitudes associated with trust in task automation. The results show evidence of performance compensation without awareness of automation failure. Trust was associated with attitudes to computer performance; task performance was associated with ratings of understanding. In the third study, a task requiring directional responses to a multi-modal display of situational information was used to elicit personal constructs associated with the cognitive compatibility (CC) of the task. Constructs were elicited using the Repertory Grid procedure. Analysis of subjective ratings of the construct dimensions indicate the multi-dimensional structure of the constructs associated with CC namely: ease or difficulty of reasoning and understanding; depth of processing or stimulus-response compatibility; learning, and experience or schema compatibility. The development of tools for the subjective measurement and prediction of SA and cognitive compatibility is discussed.

2. SITUATIONAL AWARENESS

2.1 Operational Problems

SA refers to the pilot's knowledge of the flight and tactical environment. Loss of SA plays a significant role in aircraft accidents, and it is a limiting factor on mission effectiveness. Accurate situation assessment is essential for effective planning, decision-making, and action. Reactive planning depends on accurate assessment of changing situations. SA is particularly important in "hyper dynamic" natural environments requiring fast anticipatory responses. In the military aircraft combat environment, the tactical situation can change rapidly and unpredictably. The ability to anticipate events, to plan reactively, and to predict outcomes can provide the combat "winning edge"; SA determines "who shoots and who chutes". Evidence is accumulating that under time pressure, in emergencies, experts rapidly classify situations in terms of recognised categories of experience, before selecting decision actions, dubbed as "recognition-primed decision making" (1). Problems seem to be associated with the aircraft system presenting a poor picture of the situation, either because of the increasing complexity aircraft systems, and the effects of high workload, or because of increasing cockpit automation and poor cockpit information. Military pilots complain of being "swamped with data and starved of information". On advanced highly automated civil flight decks, over-trust in automation, complacency and inattention lead to loss of SA (2). Many technical developments in aircraft cockpits are aimed at improving SA. These include helmet-mounted displays (HMD's), large area panoramic head-down displays, and Artificial Intelligence (AI) and Knowledge Based Systems (KBS) situation assessment and planning decision-aids. Aircrew training and crew resources management (CRM) packages focus on enhancing individual and team SA. The problem has now become one of predicting and evaluating what works best.

2.2 Measurement and Modelling

Theories of SA stress the importance of the pilot's continuously updated, mental representation or cognitive model of the situation, affected by limitations on attention and associated with working memory, and by knowledge of critical features and important relationships stored in semantic and episodic memory as schema and scripts. Baddeley has argued that conscious awareness is a means of co-ordinating information from a number of sources, including the present, specific episodes from the past, and projections as to the future, using a system operating through working memory (3). Testing theories requires measurement. Measurement is needed for systematic improvement of human performance, either by training, or by systems design. The ability to reliably measure, model and predict SA would be a significant development. But measurement presents practical and theoretical difficulties, affecting the validity and reliability of the data. SA is an unobservable cognitive state, not directly available for analysis. Alternative approaches include performance-based metrics, physiological indices, memory probe measures of SA knowledge, and subjective, self-report ratings. Recently, Gardiner and Java (4) have argued for the utility of subjective, experiential measures, compared with conventional measures of accuracy and performance, in distinguishing between the different states of awareness involved in conscious recollection (explicit "remember" response) compared with only having feelings of familiarity (implicit "know" response), and associated memory systems.

Following the example of workload measurement (i.e. SWAT, NASA TLX), our research efforts were directed at developing a subjective rating technique for estimating SA which would be easy to implement in the field, when performance and other measures are difficult to obtain. The resultant tool, known as SART (Situational Awareness Rating Technique), was based on aircrew constructs for SA, elicited using the Repertory Grid technique, and analysed using principal components statistical methods (5). SART was derived from the following working definition of SA:

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

"Situational awareness is the knowledge, cognition and anticipation of events factors and variables affecting the safe, expedient and effective conduct of the mission".

SART provides multi-dimensional SA rating scales. There are three primary SART rating dimensions, namely *Demand* on attentional resources (D), *Supply* of attentional resources (S), and *Understanding* (U). This simplified form of the tool is refereed to as the 3-D SART. These correspond to the 3 clusters or domains of the original aircrew constructs. The original constructs provide 10 secondary rating dimensions nested within the three primary domains (10-D SART). Ratings are obtained using 7-point Likert scales labelled Low (score 1) to High (score 7), or alternatively by marking 10cm lines. Validation studies have indicated predictive and diagnostic ability on a range of tasks, and sensitivity to task variables (6,7,8). The SART construct dimensions are summarised below:

Demand on Attentional Resources
- *Instability* of the situation i.e. likeliness to change suddenly.
- *Complexity* of the situation i.e. degree of complication.
- *Variability* of the situation i.e. number of variables and factors changing.

Supply of Attentional Resources
- *Arousal* i.e. degree of alertness or readiness for action
- *Concentration* of attention i.e. degree to which thoughts are brought to bear.
- *Division* of attention i.e. distribution, spread of focus.
- *Spare mental capacity* i.e. mental ability available for new variables.

Understanding of the situation
- *Information quantity* i.e. amount of knowledge received and understood.
- *Information quality* i.e. goodness or value of knowledge communicated.
- *Familiarity* with the situation i.e. degree of prior experience and knowledge.

More recently, a method was proposed for deriving a unitary estimate of SA from SART ratings, which would have practical utility, and which retained the characteristics of validity, sensitivity, and diagnostic power embodied in the individual SART dimensions. A unitary index can be obtained by combining the rating means, using the following simple algorithm:

SA (Calculated) = Understanding - (Demand - Supply)
for 3-D SART: $SA(c) = U - (D - S)$
for 10-D SART: $SA(c) = \Sigma U/N_u - (\Sigma D/N_u - \Sigma S/N_u)$.

This is a highly simplified model of the process whereby SA is created. It is based on *a priori* theoretical considerations, rather than statistical or empirical evaluation. Thus, caution should be exercised in the amount of weight attributed to it. *Post hoc* validation comparing the derived or calculated SA measure with experimental results has indicated mixed evidence for the utility of this approach. In a study of bi-modal cockpit warnings, SA(c) estimates from 3-D SART were found to reflect the performance data to some extent, although not fully (9). In a USAF flight simulator study of tactical operations, SA(c) estimates were calculated from 10-D SART ratings. SA(c) estimates were sensitive to experimental manipulations, and reflected the performance data. SA(c) produced a more sensitive metric than memory probes. It was judged as possibly the preferred single metric, but it was noted that the pattern of sensitivity of 10-D SART offered diagnostic value (10,11).

3. SA PREDICTION STUDY

A recent experiment was conducted to test the generality of earlier findings, and to provide a more stringent *a priori* validation of the unitary calculated SA estimate (12). The experiment used the computer-based Air Traffic Control (ATC) Simulation task provided by DCIEM, Canada under the auspices of TTCP UTP-7. ATC Simulation offers control of task demands, with task performance measures, for testing of the predictive ability of SA(c). The *a priori* prediction was that performance would improve with increasing SA(c).

3.1 Experimental Method

3.1.1 *Subjects*

Twelve non-aircrew subjects participated in the experiment. All subjects were staff at RAF IAM, between 19 and 30 years of age. Subjects were naive to both the simulated and real world ATC tasks to eliminate positive or negative training transfer effects. All were regular computer game players, to ease learning of the experimental task. All received the same amount of practice on the task.

3.1.2 *Task*

The task was provided by the ATC Simulation, Version 3.0 software programme, presented on an Apple Macintosh LC Series computer. Aircraft entered the display screen from predetermined headings and positions, at fixed "irregular" intervals, distributed to give a gradual build up of aircraft on the screen early in the scenario. The subject's task was to control the aircraft to exit the area safely, avoiding conflict, along exit path headings and at altitudes in accordance with a given screen schedule. There were no take-offs or landings. The display screen format is shown in Figure 1. Training progressed from passive viewing of a simple 3 aircraft scenario with concurrent explanation of the screen and mouse control facilities, followed by active performance of the scenario with assistance, and finally to active performance with no aiding to successful task completion.

3.1.3 *Design*

Subjects were presented with three experimental scenarios, containing 3,6, or 9 aircraft for control, selected to provide a range of task demand. Each scenario was of 5 minutes duration. The order in which the scenarios were presented was balanced to prevent order/sequence effects.

3.1.4 *Dependent Variables*

During each 5 minute scenario, SART ratings were taken three times. The scenario was paused and the 3-D SART was administered after 1 minute 40 seconds, and again after a further 1 minute 40 second period. Simple ratings of SA(r) were obtained at the same time. At the end of the scenario a 14-D SART scale was administered, comprising the 3-D and 10-D dimensions, plus a simple rating of SA(r). Performance was scored according to the number of aircraft controlled correctly, defined as leaving the area at the required altitude (1

point) and direction (1 point), converted to a percentage correct, and the number of conflicts, comprising collisions and near misses. For the purposes of the experiment, near misses were defined as separations of one screen unit (dot) space, or 1000 feet altitude.

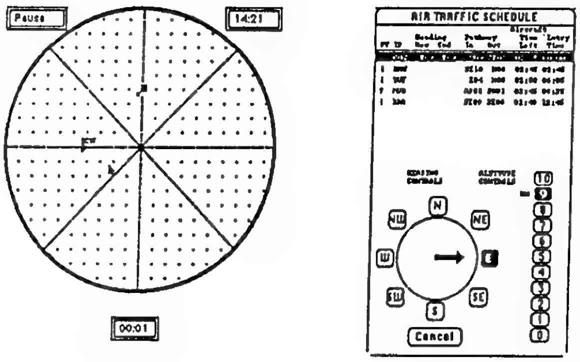


Figure 1. ATC Simulation Screen Format

3.2 Results

3.2.1 *Analysis of Variance*

Box and Cox method showed that the data met normal distribution assumptions without transformation. Analysis of the performance data and of the ratings was by ANOVA across the 3 scenario conditions. Learning and order effects were small. ANOVAs showed significant effects on the performance measures of %Correct ($F= 246.79$, df 2,22 ; $p< 0.001$) and Conflicts ($F= 8.75$, df 2,22; $p<0.01$). Post hoc tests showed the 3 aircraft condition to have more %Correct scores ($p<0.001$) and fewer Conflicts ($p<0.005$) than the other two conditions, but there were no significant differences on these measures between the 6 and 9 aircraft conditions. This strong conditions effect between the 3 and 6/9 aircraft conditions was repeated in the ratings data, including the SA(r) ratings, the individual 3-D SART ratings, and the SA(c) estimates calculated from both the 3-D and 10-D SART data ($p<0.01$ or better). Only three of the 10-D SART ratings (Information Quantity, Information Quality, Familiarity) showed no conditions effect. Significant differences between ratings for the 6 and 9 aircraft conditions were obtained for 3-D Demand ($p<0.01$), and for 10-D Variability ($p<0.01$) and Spare Mental Capacity ($p<0.05$).

3.2.2 *Correlations*

In an attempt to ascertain the common variance being accounted for by the different dependent variables, canonical correlations of the individual measures were carried out. The correlations in Table 1 summarise the main findings. Correlations greater than 0.5 are significant at the 5% level. The SA(c) estimates showed no reduction in prediction of performance compared with the simple SA(r) ratings, whilst providing greater diagnostic power. Better general predictions of performance were obtained from the 3-D SA(c) and SA(r) ratings for %Correct than for Conflicts. 3-D SA(c) provided the best prediction of %Correct performance, accounting for 40% of the variance in the data. However, the additional dimensions contributing to the 10-D SA(c) improved the prediction of Conflicts. Figures 2 and 3 show the relationships between 3-D SA(c) and performance.

	% CORRECT	CONFLICTS
SA (r)	0.581	-0.346
SA (c) 3-D	0.640	-0.468
SA (c) 10-D	0.526	-0.548

Table 1 Correlations between dependent variables

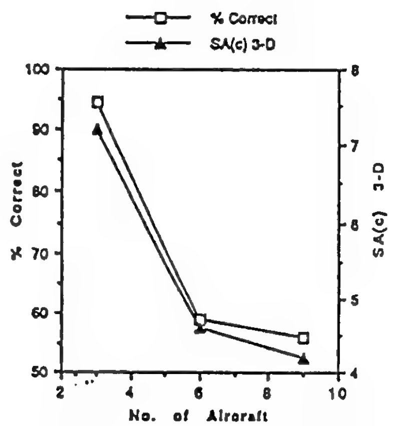


Figure 2 3-D SA(c) and %Correct

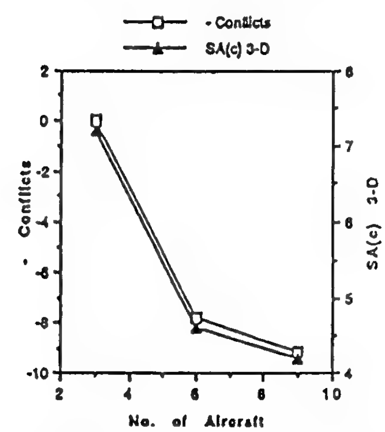


Figure 3 3-D SA(c) and Conflicts

3.3 Discussion

The task provides evidence of the effects of direct manipulation of task demands. The evidence supports the validity of *a priori* predictions of task performance, based on a unitary calculated measure of SA. Since the SA(c) formula was based on earlier findings, and not merely a best fit to the current data, the test is a more stringent assessment of the validity of the measure than the previous *post hoc* analysis.

SART was not developed to provide a unitary measure of SA. The initial aircrew knowledge elicitation presented SA as a multi-dimensional construct. Simple, uni-dimensional subjective ratings lack diagnostic power, and experimental evidence has indicated poor sensitivity and predictive performance for simple SA ratings. SA concerns knowledge, and the quality of SA is likely to be complex, and to be task, situation and operator specific. In the current study, the individual SART dimensions exhibited differential sensitivity

to the experimental conditions, and showed differences between the 6 and 9 aircraft conditions not exhibited in the performance data, nor in the SA(c) and SA(r) scores. Thus, sensitivity of multiple ratings can provide diagnostic information when simple ratings and performance measures are unaffected.

The validity of using the combined SA(c) estimates needs careful evaluation. By combining the SART scores, the assumption is made that SA can be usefully represented as a uni-dimensional concept. This is since the single SA(c) estimates represent a scalar as opposed to the 3-D vector quantities of the separate 3-D SART scores. The outcome is analogous to deriving a single measure of colour discrimination ability when colour vision is most usefully considered in terms of the ability to discriminate differences in 3-D colour space. Arguably, the required metric of SA needs to discriminate differences between situations that are important for decision making effectiveness, rather than to provide a simple quantitative index of SA.

The SA(c) formula derives from the proposition that SA is principally concerned with knowledge of the important relationships, and of the status, of variables in the situation. It is considered that SART Understanding ratings reflect knowledge of important relationships between situation variables, which largely determines SA. The ratings of SART Demand and Supply indicate the matching of attentional resources to changes in the situation variables. This attentional matching provides information on the current status of the variables. It acts as a modifier of SA, independent of knowledge of the important relationships, providing refinement and updating of the situation model in accordance with the changing status of the variables. Attentional matching increases SA when the available resources are sufficient (S>D), and reduces SA when the resources are insufficient (D>S).

This formulation is highly simplistic in contrast with complex multi-variable information processing models of human cognition, but it merely seeks to provide the best estimate of SA from the three SART scales. The simple mathematical treatment of the means of the SART ratings contrasts with the complex conjoint analysis procedures used to combine the three SWAT workload dimensions into a single measure, and the weighting of NASA TLX workload dimensions. The psychometric properties of these scales are such that SWAT conjoint analysis provides sensitivity to individual differences, and TLX weighting provides better general prediction of experienced workload (13). However, at present, it seems that the simple SA(c) formula is sufficient for the intended general predictions of SA. In most practical field work, there is merit in simplicity and ease of implementation. Refinement and additional complexity in the SA(c) model should only be introduced if improved predictions and sensitivity are needed, and if different procedures are shown to be beneficial.

4. TRUST

4.1 Joint Cognitive Systems

Aircraft systems design can help or hinder the ability of aircrew to keep themselves in the picture, and to stay ahead of the situation. Advanced, highly automated systems involve increasingly high levels of human-computer functional integration. In the future, aircrew may not be the only system cognitive resource. Rapid developments in Artificial Intelligence (AI) and Knowledge-Based System (KBS) computing technology, make it seem increasingly likely that cognitive functions will be shared between human operators and computers, in what can perhaps best be described as *joint cognitive systems*. Sharing functions and tasks with computers requires trust, but risks loss of aircrew understanding of system and mission status, and reduced aircrew SA. Aircrew SA will need to be maintained to retain ultimate authority and control. Consequently, concepts such as Human-Electronic Crew (H-EC) teamwork, co-operative functioning, and adaptive aiding are being proposed to characterise the required functional relationship between the human and computer system components. In such joint cognitive systems, trust coupled with awareness, seem likely to be the *psychological glue* which holds together the functioning of the system components.

4.2 Trust and Performance

It seems to be a truism that people generally distrust computers. Trust between humans is engendered by continuous, repetitive, and reciprocating actions. Perhaps, in the same way, it is plausible that trust will build-up when computer performance conforms consistently and predictably to expectations, in accordance with agreed goals. Investigations of the quality of teamwork in RAF aircraft tactical missions show that trust was a significant factor in distinguishing between good and poor teamwork performance (14). Trust was rated at a significantly lower level in single-seat RAF Harrier operations (i.e. human-computer teamwork) than in two-seat RAF Tornado aircraft tactical operations (i.e. both human-human and human-computer teamwork). Experimental evidence has verified that unexpected automation failure leads to a breakdown of trust, and to difficulty in the recovery of trust with a loss of faith in future teamwork performance (15,16). As trust declines, manual intervention increases. Other research has investigated how when workload is increased, over-trust or complacency develops with automatic systems (2). Complacency, coupled with vigilance problems, is likely to lead to failure to detect performance deviations and decrements in automation performance.

4.3 Associated Factors

Understanding the factors that affect trust could help design safeguards. We have reported an investigation of trust in two-seat RAF Tornado aircraft tactical operations (17). Tactical decision-making scenarios were elicited and rated for the importance of factors associated with trust in the events described. These subjective ratings showed that the demand for trust was associated with the perceived risk and the probability of negative consequences, whereas the supply of trust was related to the requirement for judgement and awareness, and the uncertainty and doubt in making the decisions. Thus, relying on others to make risky decisions calls for a large amount of trust. But if the decision requires another person exercising a high degree of awareness and judgement, and there is much uncertainty and doubt in the decision provided, then the actual trust engineered by the decision will be low. Riley has been proposed a model of the relationships between trust, operator skill level, task complexity, workload, risk, self-confidence, and EC reliability (18). Subsequent studies, in which workload and automation reliability were varied, led to refinement of the model to include the factors of fatigue and

learning about system states (19). Other research has shown how trust can be modelled as a function of parameters such as recent performance, and the presence and magnitude of a fault (20). Intervention and automation use are influenced by the combination of trust and self-confidence in operators' abilities to perform the task by manual control. Operators will allow automation to have control if they trust it; and they will take control themselves if they distrust it, providing self-confidence is sufficiently great. However, high self- confidence often produces a bias in favour of manual control.

4.4 Risks

Adaptive automation introduces new risks for successful system functioning, and along with it, the need for safeguards. Dynamic task and function allocation, with a manual default allows the possibility of unnecessary and inefficient manual intervention. With adaptive automation, different roles and responsibilities may be assigned for the same tasks at different times depending on the particular automation strategy being invoked. But this variability could easily lead to an appearance of total unpredictability, unless care is taken in the design and implementation of adaptive automation. Dynamic allocation of tasks only makes sense *if all the performers are aware of what each other is doing* (i.e. both human and computer task awareness). Otherwise, tasks might be overlooked or task contention might occur.

Sharing tasks and functions between humans and the computer introduces the risk of over-reliance and dependency on computer aiding. Reliance and dependency lead to reduced system awareness and degradation of manual skills. This becomes a problem in the event of automation failure requiring manual intervention as the default. In dynamic systems, when the information relevant for decision making changes over time, and is not static, a dynamic internal model of the task is needed to guide decision-making. An appropriate dynamic internal model of the important changing relationships will be difficult to maintain for regaining manual control following automation (21). Another problem is that operator detection of automation failure is substantially degraded with a static allocation fixed over a period of time (22). Monitoring automation performance for failures is inefficient, due in part to a natural tendency towards complacency, and because of the difficulty of maintaining vigilance without active involvement in the task. Because of problems with failure detection and manual skill degradation, manual task reallocation has been proposed as a countermeasure to monitoring inefficiency and complacency. It has been shown that short periods of intermittent manual task reallocation, or cycling between manual and automation control, reduces failures of monitoring (23). By maintaining manual skill levels, and enhancing situational awareness, manual task re-allocation helps in the event that intervention is needed following automation failure.

5. ADAPTATION FAILURE EXPERIMENT

5.1 Aim

An experiment was recently conducted at DRA CHS to investigate the effects of variations in automation adaptation performance on operator performance, and on attitudes that might affect automation use in a joint cognitive system (24). The intention was to simulate a situation involving manual task reallocation in which adaptive aiding failed, in order to measure the effect on trust in automation. Sources of failures can arise from breakdowns at different levels of system functioning. The study sought to develop understanding of how humans might react to, and cope with, high level functional failures when working with co-operative systems. In particular, the intention was to discover the sensitivity to automation adaptation failure of task performance measures and of subjective rating scale dimensions associated with trust and awareness, and to examine the structure of the relationships between these dependent variables.

5.2 Experimental Method

5.2.1 *Subjects*

Twelve non-aircrew subjects participated in the experiment. All were staff at DRA CHS.

5.2.2 *Task*

The Multi Attribute Task (MAT) battery developed by Comstock and Arnegard (25) provided the task environment. The MAT battery comprises a computer simulation of three tasks, namely tracking, monitoring and resource management, presented simultaneously. A diagram of the screen format is shown in Figure 4.

The tracking task was a two dimensional compensatory tracking task that required subjects to keep a target in the centre of the tracking window. The monitoring task required subjects to monitor and correct deviations on four gauges. The resource management task was complex, and required subjects to maintain the fuel level of two main tanks at a specified level by transferring fuel from several supply tanks. All three tasks could be operated manually, via a joy stick and keyboard, and both the system monitoring and resource management tasks also could be operated either aided or fully automatic. In the aided mode, parts of the tasks were automated, leaving the subjects to monitor the automation performance and to complete the tasks. The tracking task was always manually operated.

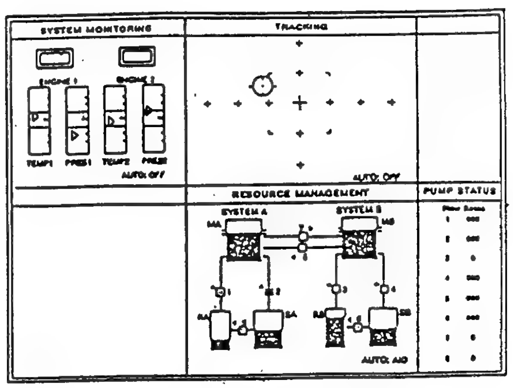


Figure 4. MAT Screen Format

5.2.3 *Scenarios*

On the experimental trials, subjects were presented with four scenarios. The independent variable was the manipulation of the *"co-operation"* given by the adaptive aiding, i.e. the extent to which the automation performed according to expectations. In all four scenarios, the *invocation* of the automation occurred

automatically, i.e. changes in the automation level were initiated by the computer, and not by the operator/subject. Co-operation was experimentally manipulated by providing two co-operative and two unco-operative scenarios. In all four scenarios, the frequency of events requiring action, and the resultant task demands, increased as the scenarios progressed. In the co-operative scenarios, the level of aiding provided by the automation increased appropriately with the event frequency and level of task demand. In one co-operative scenario, the system monitoring task went aided and remained so; the resource management went aided, and then fully automated (*Co-op 1*). In a second co-operative scenario, the resource management task went aided, and remained so; the system monitoring went aided, and then fully automated (*Co-op 2*). In the unco-operative scenarios, the level of aiding initially increased appropriately, but then shortly after the onset of a period of particularly high event frequency and task demand, when the level of aiding could be expected to be increased or at least maintained, the aiding automatically re-allocated to manual. In one unco-operative scenario, the system monitoring task switched to, and remained at aided; the resource management task also went aided, then warned to go fully automated, but then reverted to manual (*Unco-op 1*). In the other unco-operative scenario, the resource management task switched to, and remained at aided; the system monitoring task also went aided, then warned to go fully automated, but then reverted to manual (*Unco-op 2*). Each experimental scenario was of five minutes duration. The sequence of events is illustrated in Figure 5.

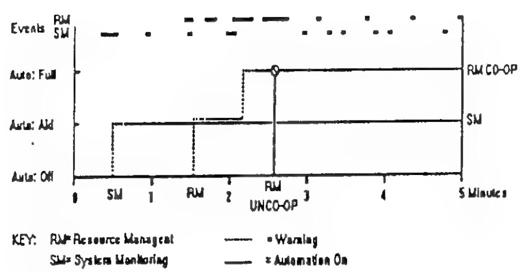


Figure 5. Event Sequence

5.2.4 *Design*

All the subjects were presented with the four experimental scenarios. Subjects received a 5 minute practice session before each experimental session, which included combinations of automation and manual operation of the three tasks. All subjects received the same amount of practice on the task, including manual, aided and fully automated modes. Subjects were instructed that the computer would attempt to give the appropriate level of aiding, but that accuracy of the computer's judgement was imperfect and that compensation may be required if the aiding failed to be appropriate. The order in which the scenarios were presented was balanced to prevent order / sequence effects.

5.2.5 *Dependent Variables.*

Dependent variables comprised computer measures of task performance and subjective ratings. The following MAT task variables were recorded: root mean square (RMS) error on the tracking task; the number of correct resets, incorrect resets, and mean reset time on the system monitoring task; tank 1 deviations, tank 2 deviations, and the number of pump activations on the resources management task. After each scenario, subjects provided subjective assessments on 7-point Likert rating scales of the timeliness and appropriateness of the computer aiding on dimensions (low to high) of 17 constructs related to trust and awareness. The constructs for ratings were defined as follows:

Confidence - Confidence in own ability to successfully complete the tasks with the aid of the adaptive automation.
Self Confidence - Confidence in own ability to successfully complete the tasks.
Accuracy - Accuracy of own performance on the tasks with the aid of the adaptive automation.
Self Accuracy - Accuracy of own performance on tasks.
Automation Confidence-Confidence in ability of the machine to support successful completion of the tasks.
Automation Accuracy - Accuracy of machine in supporting successful completion of tasks. *Automation Dependability* - To what extent can you count on the machine to provide the appropriate support to the tasks.
Automation Reliability - To what extent can you rely on the machine to consistently support the tasks.
Automation Predictability - Extent to which you can anticipate and expect the machine to support the tasks.
Risk - The probability of negative consequences of relying on the machine to support successful completion of the tasks.
Impact/Survivability - The severity and criticality of adverse or negative consequences of relying on the machine to support successful completion of the tasks.
Decision Complexity - The extent to which the machines' decision on when and how to intervene and support the task can be regarded as a simple and obvious choice.
Uncertainty/Doubt - The extent to which you have confidence in the machines' decision on when and how to intervene and support the task.
Judgement/Awareness - The extent to which the machines decision on when and how to intervene and support the task requires assessment, knowledge, and understanding of the task.
Faith - To what extent you believe that the machine will be able to intervene and support the tasks in other systems states in the future.
Demand for Trust - Level of trust required from you when the machine intervenes and supports the task.
Supply of Trust - Level of trust actually provided by you when the machine intervenes and supports task.

In addition to the above, situational awareness (SA) was rated using 3-D SART dimensions described earlier, i.e. Demand, Supply, and Understanding.

5.3 Results

5.3.1 *Analysis of Variance.*

ANOVAs on the dependent variables showed no clear pattern of effects arising from the manipulation of computer co-operation. Significant differences were found between the scenarios on Resource Management tank level 1 and 2 deviations ($p<0.001$). Newman Keuls tests showed significantly more deviations in the Co-op2 scenario than in the other three scenarios ($p<0.01$). There was a small but significant difference between the subject groups on ratings of the supply of trust ($p<0.05$). There were no significant subject

group/scenario condition interactions, and thus no proof of order or transfer effects between the scenario conditions.

5.3.2 *Correlations.*
Correlation analysis was performed on the performance and ratings data. This analysis revealed significant correlations between the many of the variables. A schematic representation of the significant correlations is provided in Figure 6, following the style used by Riley (18,19). In Figure 6, variables with significant correlations (r > 0.40) are linked by lines, with the strength of association indicated by the line width.

5.3.3 *Factor Analysis.*
Factor analysis of the subjective ratings found that four factors accounted for 62% of the total variance in the data. The results are summarised in Table 2, with ratings variables that obtained significant loadings on the four factors (> 0.45) shown in order of reducing weight, with positive or negative values (+/-ve).

Factor 1 (-ve) 21.55 % Variance.	Factor 2 (-ve) 17.65 % Variance.	Factor 3 (+ve) 11.97 % Variance	Factor 4 (-ve) 10.95 % Variance
Auto. Reliability Auto. Confidence Auto. Dependability Auto. Accuracy Auto. Predictability Supply of Attentional Resources Supply of Trust (+ve)	Self Accuracy Confidence SA(c) Self Confidence Accuracy Understanding of Situation	Impact (-ve) Supply of Trust Supply of Attentional Resources Demand for Trust	Uncertainty / Doubt Faith Decision Complexity Demands on Attentional Resources

Table 2. Factor Analysis of Ratings

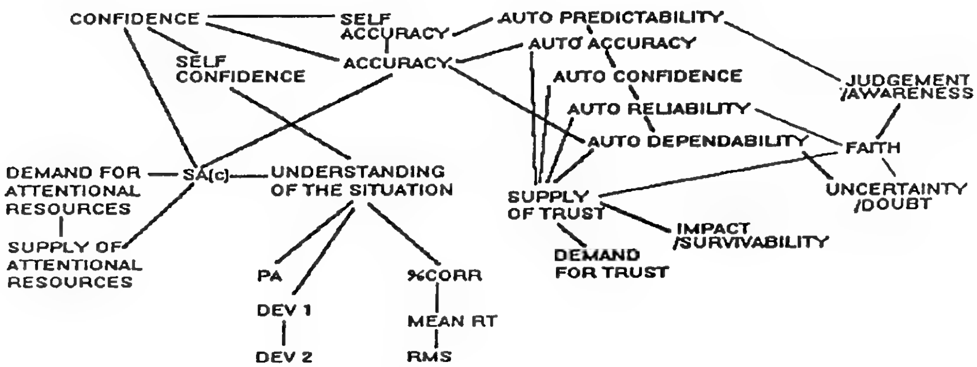


Figure 6. Schematic Structure of the Correlated Variables

5.4 **Interpretation**
Analysis of the results by ANOVA indicated that the subjects successfully compensated for the variability in the computer aiding performance, but with little effect on their assessments and attitudes regarding the automation. Despite clear instructions to monitor the imperfect aiding invocation, with regard to timeliness and the level of aiding provided, most subjects seemed unaware of the experimental manipulation of computer co-operation. Only 3 subjects noticed the computer warning to go fully automated, and then failing to do so. Trust ratings varied about the middle of the rating scale (mean = 4.23; SD = 1.15). Mean trust supply ratings were higher, and SA(c) means were lower, following the co-operation scenarios, but again the differences did not achieve statistical significance.

The correlation analysis and factor analysis show the structure of the relationships between the variables. Supply of trust was related to confidence in automation performance and to its perceived accuracy, reliability, and dependability. An associated automation performance factor in the factor analysis (Factor 1) accounted for the largest proportion (21.55%) of the variance in the ratings data. The correlations show trust was inversely associated with impact / survivability, or the negative consequences of relying on computer to support the task, i.e. the more adverse the consequences were perceived to be, the less trust was supplied. A similar trust related factor (Factor 4) accounted for 11.97% of the ratings variance. Faith, referring to future performance, was more associated with the requirement for judgement and awareness in the computers decisions, with weak associations with decision uncertainty and doubt, and with perceived automation reliability. A similar, relatively weak uncertainty/faith related factor (Factor 4) accounted for 10.95% of the ratings variance. Self-confidence was linked to assessments of the accuracy of own performance, and to awareness and understanding of the situation. Self confidence / accuracy was associated with a strong factor (Factor 2), which accounted for 17.65% of the ratings data. SART Understanding of the situation was associated with the MAT performance data. Thus, consistent with the evidence of successful compensation, performance on the task was associated more with awareness and understanding of the task, which allowed the compensation to occur, and less with the perceived performance of the automation.

Several points follow from the above. It seems likely that subjects experienced difficulty in distinguishing between tasks

demands arising from the external situation, and demands associated with variability in the performance of the adaptive aiding. The subjective assessments were surprisingly uncritical of the unco-operation scenarios. There was no substantial loss or gain in trust, and consequently no basis for assessing trust recovery following unco-operative scenarios. Equally, there was no evidence of misplaced trust or of attitude changes to resolve cognitive *dissonance* arising from any mismatch between attitudes and behaviour. The functioning of the automation generally seems to have been compatible with subjects expectations. But then initial levels of trust would need to be high for awareness of an inappropriate reallocation to cause a revision of attitudes. It seems likely that the appropriateness of aiding invocation, level of assistance, and reallocation strategy are difficult to judge in a dynamic situation. The demands associated with the automation performance were communicated through a common interface, and shared a common representation with the external task demands. Variations in the task due to the automation could be distinguished only by inference from differences between expected and actual automation performance. It is possible that the expectations for automation performance were not clearly and rigidly formulated, and that the compensation occurred smoothly and naturally, without a reluctant hand-over. Subjects appear to have maintained throughout a reasonably high level of confidence in their own ability, and in the computer's assistance, to perform the task. While self confidence is maintained, and compensation occurs without awareness, deviations are likely to be attributed to own, rather than computer, performance. Confidence in adaptive aiding seems dissociated from adaptation performance. Sen and Boe (26) report a similar dissociation between confidence and accuracy in computer-aided decision making. Confidence may not be a reliable predictor of joint cognitive system performance.

Safeguards are needed against breakdown or failure in performance to ensure that operator trust in system functioning is maintained at realistically appropriate levels, without adversely affecting situational awareness. The "*First Law of Adaptive Aiding*", that computers should be able to take tasks, but not give them, is supported by these findings (23). Automatic re-allocation of tasks to manual seems close to a violation of this First Law. Careful consideration of the procedures is needed for implementation of dynamic task allocation and re-allocation. Such variable assistance and allocation could lead to unacceptable unpredictability. Awareness of the current task allocation strategy is an important factor for system effectiveness, but this may not easily be achieved with *seamless* adaptive aiding. Careful consideration needs to be given to the design of the functional interface in joint cognitive systems to ensure that the appropriate level of awareness of the current task allocation is easily maintained. Awareness is needed to avoid task contention, and to ensure that tasks are not overlooked or performed incorrectly. Trust is built on awareness of proven performance. Bi-directional compensation without awareness might engender false trust by others outside of "the team".

Conflicting or unco-operative intelligence could arise from poor design, such as inappropriate adaptive logic. For this reason, adaptive strategies, such as manual reallocation, will need careful adaptive logic to ensure their appropriateness. Alternative forms of adaptive logic have been proposed based on either critical events, pilot performance measurement, pilot physiological assessment, and pilot modelling. The validity of the threshold criteria for triggering allocation and re-allocation will be critical. Predictions of pilot workload would seem to be the logical candidate for a model-based adaptive logic, particularly for manual task re-allocation. However, sufficiently reliable individual workload threshold criteria will be difficult to obtain from currently available generic workload models and measurement tools. Given the low predictive validity of human performance and workload models, the possibility of an inappropriate re-allocation from an operator model-based logic, or any other logic, will need to be anticipated in the design of adaptive systems. Measures will need to be taken to guard against the consequences of inappropriate allocation, adaptation breakdown or failure, or of what might otherwise appear to be unco-operative automation.

As a safeguard, the system will need to establish operator willingness and readiness to accept tasks before reallocation. In addition, safeguards will need to be built into the human-computer interface to ensure that the operator has the necessary awareness and control of the current functional and task configuration. It may not be sufficient to provide legends for automation status; pictorial representations or dialogue may be needed for comparing the pilot's expectations with computer's plans and intentions. Poor co-operation could lead to mistrust, or be perceived as systematic and intentional, engendering a sense of paranoia. Safeguards may be needed to ensure that there is a shared common understanding of all the system meta- and sub-goals underlying co-operative functioning, i.e. the prime directives.

Ultimately, the degree of trust engendered in any cognitive system can be considered to be related to degree of agreement between the expectations for system performance, and the performance perceived to have been achieved. This agreement can be regarded as a matching and consistency between actual and intended goals, functions and tasks of the system components. A high level of agreement will be required between the human and computer cognitive components and functions for effective joint cognitive systems.

7. COMPATIBILITY.

7.1 Understanding Understanding

Both the foregoing studies provide evidence of the central role of *understanding* in the association between situational awareness and task performance. Understanding, when assessed using SART, draws upon the ideas of information, knowledge, and experience or expertise, which in systems design fall into the realms of *cognitive engineering*. This enters an area of system cognitive requirements that have traditionally been regarded as issues of consistency or matching of design with human expectations, perhaps best referred to as *cognitive compatibility* (CC).

7.2 Compatibility Types

Concern with compatibility in systems design originates from the ideas of Arnold Small on stimulus-response compatibility, presented to the first meeting of the Ergonomics Society in 1951 (27), and the subsequent work of Fitts and Seeger (28).

McCormick and Sanders (29) define compatibility in relation to human engineering design,as follows:

" the spatial, movement, or conceptual relationships of stimuli and responses, individually or in combination, which are consistent with human expectations."

The authors go on to describe a taxonomy of different types of compatibility. *Conceptual compatibility* is described as referring to conceptual associations, *intrinsic* in the use of codes and symbols, or culturally acquired associations.

7.3 Mental Models

More recently, Wickens (30) has discussed the importance of compatibility (or *congruence*) between levels of representation which form the basis for understanding systems, namely the physical system (which is analog), the *internal representation* or "*mental model*" (which should be analog), and the interface between the two. Compatibility between the real system and the mental representation he argues is clearly a matter of training. If both the both the physical system and the mental representation are analog, as they should be for correct understanding, then it is important that the display should be formatted in a way which is compatible with the other two. In a comparison of the relative importance for design of consistency and compatibility, Andre and Wickens (31) use the notations S for stimulus, C for *comprehension* or *cognitive understanding*, and R for response, suggesting that S-C *mappings* are concerns of *cognitive compatibility* and S-R mappings are concerns of *physical compatibility*. Pictorial formats for aircraft, and schema-based display formats for improved SA are practical examples of how these ideas can be applied in systems design (32).

7.4 Intuitive Interfaces

The research literature on human-computer interaction and usability provides guidance for design of *intuitively* useful features, similar to ideas of cognitive compatibility. Dix et al (33) report that intuitive features of graphical user interfaces are considered to include WYSIWIG (what you see is what you get), and simplicity of mapping between representation and product. Usability principles for *learnability* include predictability, sythesizability, familiarity, generalisability, and consistency.

7.5 Measurement

There is a substantial body of theory concerning CC, but a lack of measurement tools and systematic design procedures, protocols and guidance. As noted earlier in the work on SA, the testing of theories requires measurement. Measurement also is needed for systematic improvement of human performance, either by training, or by systems design. Methods have been proposed for measuring the cognitive complexity of displays based on network analysis (34, 35). Generally, the quality of CC has to be inferred indirectly from objective measures performance, whilst making assumptions about the underlying and hidden cognitive structures and processes. New *cognitive measures* are needed to provide sensitivity and diagnostic power. The problem with measuring cognition is that it is not directly observable. Subjective measures provide one way forward, but they pose important theoretical and methodological problems with regard to the issues of measurement validity and reliability. However, experience with subjective workload measures and SART indicate that validated subjective techniques can have practical utility and predictive power. This final section describes an ongoing DRA CHS study aimed at developing a validated, subjective, self-report tool for measuring CC.

8. COMPATIBILITY STUDY

8.1 Aim

The aim of the study was to develop a subjective rating scale for measuring CC. The objectives include: the development of a task environment for manipulating (theoretically) relevant CC variables, namely modality, spatial, movement, and conceptual compatibility; eliciting personal constructs associated with CC from exposure to the task environment; analysing the structure of the constructs; developing a subjective rating scale based on the construct structure; validating the rating scale using performance data; supplying the validated tool in support of related CC application area research. This report concerns two initial phases of the study conducted to date, namely the process of construct elicitation, and the identification of the construct structure.

8.2 Method

8.2.1 *Subjects*

The subjects were non-aircrew members of staff at DRA CHS, aged between 20-30 years old. 30 subjects participated in the first phase of the study; 20 subjects participated in the second phase. Most subjects participated in both phases.

8.2.2 *Task*

The same task was used in both phases of the study. The task environment simulated flying an aircraft in tactical situations using a DRA CHS computer system. It involved the presentation of information on other aircraft locations and required orienting responses in relation to those locations. Spatial reasoning, visualisation, mental rotation and decision-making were key cognitive task components.

The presentation of the information was designed to provide correlated and uncorrelated task cues demonstrative of varying modality (MD), spatial (SP), movement (MV), and conceptual (CN) domain compatibility. Verbal and spatial left/right information was presented using the visual or auditory modalities (uni-modal), or simultaneously in both modalities (bi-modal). Spatial information was provided by the position of the cues on the visual display, or by the ear of auditory presentation over headphones. The stimuli were written or spoken words (i.e. "LEFT" and "RIGHT"), a non-directional visual symbol (i.e. XXXX), or a simple auditory tone. Subjects were instructed that the auditory modality presented command information (i.e. "go to the ....."), and that the visual

modality gave status information (i.e. "on the ...."). The aircraft were identified as threats (red) or targets (white). The presentation of the direction of flight of the own aircraft was varied between trials, presented as either track-up (ego-centric display) or track-down (exo-centric format). The direction was indicated on the visual display by the position and the direction of an arrow symbol. A keyboard was provided for responding with 2 keys labelled "LEFT" and "RIGHT" respectively. The orientation of the keyboard was varied between trials. The keys were oriented vertically, or arranged horizontally, with the labels LEFT and RIGHT on spatially appropriate keys, or reversed. The task required a directional (left/right) keyboard response according the status of the other aircraft, taking into account the direction of flight of the own aircraft. Subjects were instructed to move away from threat aircraft, and to close on target aircraft. Training on this initially complex task was provided with sample situations before test trials.

Through this task environment, compatible and incompatible instances were provided of the domain manipulations i.e. MD (visual/auditory), SP (left/right localisation; ego-centric/exo-centric display format), MV (vertical/horizontal keyboard) and CN (target/threat; command/status). These domain manipulations were provided in all possible 11 combinations (i.e. 2-way - MD/CN, SP/CN, SP/MD, MV/CN, MV/MD, SP/MV; 3-way - MD/SP/CN, MD/MV/CN, SP/MV/CN, SP/MV/MD; 4-way - MD/SP/MV/CN). The result was a total of 44 situations, requiring an equal number of LEFT and RIGHT responses, with an equal number of correlated and uncorrelated task cue combinations demonstrating degrees of compatibility and incompatibility.

8.3 Construct Elicitation

The first phase of the study sought to elicit personal constructs associated with CC, for use in subsequent rating scale development.

8.3.1 *Procedure*

A version of Kelly's Repertory Grid procedure, similar to that employed in the development of SART, was used to elicit personal constructs associated with CC. Subjects were given the following TTCP UTP-7 agreed working definition of CC:

"*....the facilitation of goal achievement through the display of information in a manner which is consistent with internal mental processes and knowledge, in the widest sense, including sensation, perception, thinking, conceiving, and reasoning.*"

After training, subjects were then presented with three randomly selected examples of the correlated and uncorrelated directional "LEFT" and "RIGHT" situations. Briefing information varied the perception of the stimuli as command or status, threat or target cues, and ego or exo-centric information. Subjects were asked to think about each situation and respond by pressing a left or right labelled key. Response times and errors were recorded automatically. Then the subject was required to identify two of the situations which contained something important for cognitive compatibility, in accordance with the working definition, that was not a feature of the third situation. The subject was then required to identify and describe the discriminating characteristic, and the construct thus elicited was recorded. Subjects were then presented with all 22 situations in a random order. They were asked to respond to each situation using the keyboard, and then to rate the situation on the elicited construct using a 10cm line scale labelled "low" and "high". Again, response times and errors were recorded automatically. The ratings were measured and recorded. Additional constructs were elicited using different triads of situations. A total of 56 CC construct dimensions with associated situation ratings, and performance data, were obtained in this way from 30 subjects.

8.3.2 *Analysis of Results*

14 of the constructs were elicited more than once (i.e. confusing × 5; spatial/orientation × 4 -; simple, consistent, expected/expectancy, natural × 3 ; automatic, complexity, concise, contradictory, instinctive, logic/al, obvious, difficult × 2). This gave a total of 32 unique dimensions. The frequency of errors was extremely low and insufficient for statistical analysis. An unbalanced ANOVA performed on the response time data (log transformation) considered the effects of trial number, run, left/right response, information condition (11 levels), compatibility condition (2 levels), and subjects. All the effects were demonstrated ($p < 000.1$) except for a left/right bias. Consequently, the left and right version of each condition were considered to be equivalent for purposes of further analysis. The structure of the elicited CC constructs was investigated by statistical analysis of the 56 construct / situation ratings using principal component technique, with the left/right conditions considered to be equivalent. Principal components analysis revealed that the 5 components explained most of the variance (74.98%). The majority of the dimensions (i.e. 35) were associated with the 1st component which accounted for 34.44% of the variance. Only 8 were associated with the 2nd component (10.48% variance); 7 with the 3rd component (12.35% variance); 6 with the 4th component (11.22% variance); and 4 with the 5th component (6.49% variance). Identification of the dimensions, and of the underlying structure of the constructs was unclear from this analysis. Correlations of the response times (log transformation) with the subjective ratings showed the strength of association of the construct dimensions with performance. Only 9 of the correlations for dimensions with response times failed to reach statistical significance at the 5% level.

8.3.3 *Initial Construct Selection*

Guided by this analysis, constructs were eliminated from further evaluation using the criteria of lack of association with response times, coupled with low elicitation frequency and weakness of component loading (i.e. abnormal, bias, pre-determined, reinforcement). In addition, constructs were also rejected which were task specific and not generic (i.e. spatial orientation / orientation), and which were directly descriptive of performance, and considered to be lacking of diagnostic power (i.e. easy, difficult, quickly, speed, confidence). So, of the 32 original unique constructs, 22 constructs and associated dimensions were selected for further evaluation. The 22 constructs selected in this way, with associated descriptions are listed in Table 3.

8.4 Construct Structure Identification

The number of constructs elicited in the first phase (i.e. 22) was relatively large for a practical rating scale tool. For ease of implementation, a reduced set of rating scales is needed, or a

hierarchical structure with fewer dimensions in the most simplified form, as in SART. This requires an understanding of the relationships between the constructs. So, in order to gain a better understanding of the structure of the constructs, and their relationships with task performance, a second experimental phase was undertaken.

8.4.1 *Procedure*

To simplify the experimental procedure, 16 situations were selected from the 44 situations used in the construct elicitation phase of the experiment. The selection was guided by the response time performance data obtained in the elicitation phase. The response times were used to indicate a representative range of the difficulty in the experimental conditions. Situations were selected that provided slow, fast and intermediate mean response times in the first phase of the study. These comprised compatible and incompatible instances of five of the 2-way domain manipulations (MD/CN, SP/CN, SP/MD, MV/CN, MV/MD), and three of the 3-way manipulations (MD/MV/CN, SP/MV/CN, SP/MV/MD), requiring an equal number of LEFT and RIGHT responses. After training with sample situations, subjects were presented with the 16 test situations in a random order. They were required to respond to each situation with the correct LEFT or RIGHT response. Response times and errors were recorded automatically. Once a response was given, subjects were asked to rate their experience of the situation on the 22 construct dimensions using 10 cm scales labelled "low" to "high", in the order of listing in Table 3. A written description of the construct dimensions was provided for reference during the rating task. The ratings were measured and recorded.

Construct	Dimension
1. Simplicity	Degree of simplification and ease in the situation.
2. Straightforward	Degree to which the situation is direct and clear cut.
3. Clarity	Degree to which the situation is clear and lucid.
4. Naturalness	Degree to which the situation appears normal, as in nature, and not requiring learning.
5. Conciseness	Degree to which the situation is brief and to the point.
6. Automaticity	Degree of habit and lack of conscious thought in the situation.
7. Confusability	Degree to which the situation is perplexing and bewildering.
8. Logicalness	Degree to which the situation can be characterised by coherent and valid reasoning.
9. Understandability	Degree to which the meaning of the situation is known and comprehensible.
10. Complexity	Degree of complication (number of closely connected parts) and convolution of the situation.
11. Obviousness	Degree to which the situation is self evident and plain.
12. Instinctiveness	Degree of gut feeling and innate reaction to the situation.
13. Association	Degree of mental connection of ideas, feelings, or sensations.
14. Confliction	Degree to which the situation causes opposition and contention.
15. Contradiction	Degree of antagonism and negation in the situation.
16. Representativeness	Degree to which the situation is depicted typical of the mental picture.
17. Expectation	Degree to which the situation is predictable and anticipated.
18. Recognisability	Degree to which the situation belongs to the same class as something previously seen or known.
19. Comprehensiveness	Degree to which the scope or content reinforces the situation.
20. Intuitiveness	Level of spontaneity and insightfulness in the situation.
21. Consistency	Level of agreement and accordance within the situation.
22. Familiarity	Degree of acquaintance and past experience of the situation.

Table 3. Initial Construct Dimensions

Factor 1 (-ve) 28.43% Variance	Factor 2 (+ve) 23.33% Variance	Factor 3 (-ve) 6.71% Variance	Factor 4 (-ve) 13.35% Variance	Factor 5 (-ve) 9.43% Variance
Confliction (+ve)	Representativeness	Conciseness	Familiarity	Simplicity
Contradiction (+ve)	Expectation		Recognisability	Comprehensiveness
Confusability (+ve)	Instinctiveness			
Complexity	Intuitiveness			
Straightforward	Automaticity			
Clarity	Association			
Understandability	Obviousness			
Consistency	Naturalness			
Naturalness				
Logicalness				
Obviousness				
Instinctiveness				
Automaticity				

Table 4. Factor Analysis of Ratings

8.4.2 *Analysis of Results*

The frequency of errors was low and insufficient for meaningful statistical analysis. Principal component analysis of the ratings revealed that 5 components explained most of the variance. The constructs loading on these factors, with correlation coefficients greater than 0.5, are shown in Table 4, listed in order of magnitude of correlation. Four constructs with significant loadings on Factor 1, also obtained stronger loadings on Factor 2, namely instinctiveness, automaticity, obviousness, and naturalness. After applying a sign correction, a classic cluster analysis was performed on the inter-correlations. The groupings identified are shown in Table 5, listed in terms of similarity, and illustrated graphically in the diagram in Figure 6. In this Figure 6, the length of the link lines is inversely proportional to the strength of correlation between construct ratings i.e. short link lines are indicative of synonyms (and antonyms).

Group 1	Group 2	Group 3		Group 4	Group 5
		Sub-group 3.1	Sub-group 3.2		
Simplicity	Straightforward Clarity Confusability Complexity Confliction Contradiction Logicalness Consistency Understandabil'ty	Naturalness Obviousness Instinctiveness Automaticity Representat'ness	Association Expectation Comprehens'ness Intuitiveness	Recognisability Familiarity	Conciseness

Table 5. Construct Groupings from Cluster Analysis

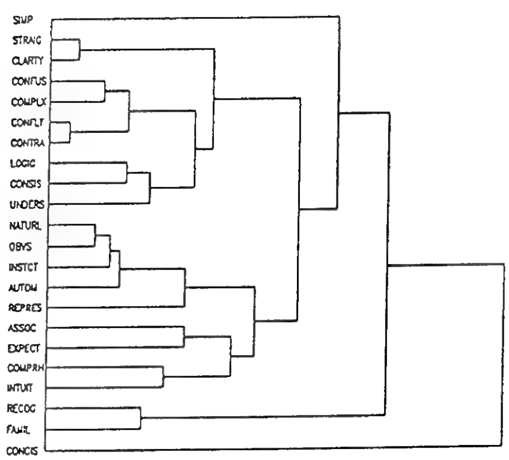


Figure 6. Structure of the Construct Clusters

Correlation coefficients were calculated between the response times (log transformation) and the ratings, taking into account differences between subjects. The correlations are listed in Table 6.

Construct	Correlation Coefficient	Statistical Significance
Automaticity	- 0.592	p< 0.001
Straightforward	- 0.588	p< 0.001
Contradiction	0.581	p< 0.001
Understandability	- 0.578	p< 0.001
Confliction	0.575	p< 0.001
Clarity	- 0.569	p< 0.001
Naturalness	- 0.568	p< 0.001
Obviousness	- 0.554	p< 0.001
Consistency	- 0.544	p< 0.001
Representativeness	- 0.523	p< 0.001
Familiarity	- 0.518	p< 0.001
Instinctiveness	- 0.516	p< 0.001
Complexity	0.508	p< 0.001
Confusability	0.500	p< 0.001
Recognisability	- 0.494	p< 0.001
Simplicity	- 0.486	p< 0.001
Intuitiveness	- 0.468	p< 0.001
Comprehensiveness	- 0.467	p< 0.001
Logicalness	- 0.442	p< 0.001
Association	- 0.440	p< 0.001
Expectation	- 0.379	p< 0.001
Conciseness	- 0.197	p< 0.050

Table 6. Correlations of Ratings with Response Times

8.5 Interpretation

Constructs associated with CC have a multi-dimensional structure. Subjective measures of CC should seek to capture, and to be sensitive to, this structure. To develop a practical rating scale, consideration needs to be given to the ease of implementation, as well as to the issues of validity and reliability. Simplification is the key to easy implementation of rating scales. A CC rating scale using all 22 elicited construct dimensions would be time consuming and cumbersome to administer.

The factor and cluster analysis provide indications of possible levels of simplification, based on the groupings and semantic associations. The strength of intercorrelations provide evidence of synonyms (e.g. Straightforward = Clarity; Confliction = Contradiction), and indicate redundant constructs that probably offer little or no added sensitivity and diagnostic power. The correlations between the ratings and response times indicate a further important basis for selection and simplification according to association with performance, sensitivity to conditions affecting performance, and hence

potential predictive power. On this evidence, for instance, Automaticity probably offers the strongest, and Conciseness the weakest, prediction of performance variance (36% and 5% respectively).

Consideration of the principal components analysis indicates how the inclusiveness of constructs associated with different factors could measure variance in CC ratings. Factors 1 and 2 account for 51.76% of the total variance; the addition of Factor 4 brings this figure up to 65.11%; adding Factors 3 and 5 give a notional maximum of 81.25 %. Measuring Factors 1, 2 and 4 could be sufficient for estimating variance in CC for many purposes. On the other hand, adding the few constructs associated with Factors 3 and 5 (i.e. 3), could provide a more differentiated characterisation of CC, with greater diagnosticity, at relatively little extra cost in terms of added complexity. Similar considerations apply to the clusterings. Three cluster groupings, namely Groups 2, 3 and 4, comprise the constructs loading on Factors 1, 2 and 4. Cluster analysis indicates two Group 3 sub-groups loading on Factor 2, and separates the two constructs loading on Factor 5, i.e. Simplicity and Comprehensiveness.

Some indication of the identify of the factors and sub-groups can be obtained from consideration of the strength of factor loadings and the semantic content of the groupings. On the basis of the present data, it seems reasonable to propose that the constructs fall into three main categories, namely:

a. Ease or difficulty of reasoning, working memory, *intellectual* and mostly S-C compatibility (high Factor 1 loadings, Group 2 cluster).
b. Depth of processing, *automatic*, and mostly S-R compatibility, with associated internal processes (high Factor 2 loadings, Group 3.1 & 3.2 clusters).
c. Learning, knowledge and experience, and *schema* compatibility (high Factor 4 loadings, Group 4 cluster).

There are similarities between this breakdown and the differences between rule, skill and knowledge-based behaviour. The intention is to use this, or a similar categorisation, as the basis for a simplified CC rating scale. This analysis goes further than SART in characterising the nature of cognition and understanding by distinguishing between levels and ease of processing. The Group 1/Factor 3 and Group 5/Factor 5 constructs could provide increased scope of rating assessment. The constituent elicited constructs of the main categorisations could provide increased depth of diagnosticity, and sources for repeated estimates of the category variance. Formal, systematic procedures are needed to verify the validity of any hypothesised classification, and of derivative measurement tools. Verification and validation will be undertaken in future phases of this work aimed at key developing schema-based display formats for HMD symbology and adaptive aiding interfaces.

5. CONCLUSIONS

The relationships between constructs as differentiated as awareness, trust and compatibility are unlikely to be simple. Principal components analysis of construct ratings provides a means of identifying important relationships between cognitive constructs with implications for systems design. SART has shown how this knowledge can be turned into a useful subjective measurement tool, with some power for predicting performance. *Understanding* seems to be a particularly important for SA and performance. Understanding seems likely to be a continuing factor in working with joint cognitive systems. Understanding is needed for assessing task and system status and functioning, and in establishing a valid basis for the assignment of trust in automation and computer aiding. Methods are needed for systematically enhancing SA and understanding in systems design. The measurement, modelling and prediction of cognitive compatibility offers a way forward. Personal constructs associated with cognitive compatibility appear to have some common structure, that has some consistency with theory. This could provide the basis for the development of protocols for subjective, self-report rating of CC associated with SART.

6. REFERENCES

1. Klein, G. "Recognition-Primed Decisions", Advances in Man-Machine Systems Research" *in* "Advances in Man-Machine Systems Research", Rouse, W (Ed), *5*, 47-92. Greenwich, CT: JAI Press.
2. Parasuraman, R., Molloy, R., and Singh, I. "Performance Consequences of Automation-Induced Complacency". Int. J. Av. Psych., *3*, 1, 1993, pp 1-23.
3. Baddeley A. " Working Memory and Conscious Awareness", *in* "Theories of Memory" Collins A.F., Gathercole, S., Conway, M, and Morris, P. (Eds), Hove: Erlbaum, 1993.
4. Gardiner J.M., and Java R.I. "Recognising and Remembering", *in* "Theories of Memory" Collins A.F., Gathercole, S., Conway, M, and Morris, P. (Eds), Hove: Erlbaum, 1993.
5. Taylor R.M. (1990), 'Situation awareness rating technique (SART): The development of a tool for aircrew systems design', in AGARD CP 478, Situation Awareness in Aerospace Operations, AGARD, Neuilly sur Seine.
6. Selcon, S.J. and Taylor R.M. 'Evaluation of the situation awareness rating technique as a tool for aircrew systems design', in AGARD CP 478, Situation Awareness in Aerospace Operations, 1990, AGARD, Neuilly sur Seine.
7. Taylor, R.M. and Selcon, S.J. , 'Cognitive quality and situational awareness with advanced aircraft attitude displays', in Proceedings of the HFS 34th Annual Meeting, 1990, *1*, pp 26-30, HFS, Santa Monica.
8. Selcon, S.J., Taylor, R.M. and Koritsas, E. 'Workload or situational awareness?: TLX vs SART for aerospace systems design evaluation', in Proceedings of the HFS 35th Annual Meeting, 1991, *1*, pp 62-66, HFS, Santa Monica.
9. Selcon, S.J., Taylor, R.M. and Shadrake, R. , 'Multi-modal cockpit warnings: Picture, words, or both?, *in* Proceedings of the HFS 36th Annual Meeting, 1, 57-61, HFS, Santa Monica. 1992.
10. Crabtree M.S., Marcelo R.A., McCoy A.L. and Vidulich, M.A. , " An examination of a subjective awareness measure during training on a tactical operations simulator", in Jenson, R. & Neumeister, D. (Eds), Proceedings of the 7th International Symposium on Aviation Psychology, OSU, Columbus. 1993.
11. Vidulich, M.A., Crabtree M.S. and McCoy, A.L. " Developing subjective and objective metrics of pilot situation awareness", *in* Jenson, R. and Neumeister, D. (Eds), Proceedings of the 7th International Symposium on Aviation Psychology, OSU, Columbus. 1993.

12. Taylor, R.M., Selcon, S.J., and Swinden, A.D. "Measurement of Situational Awareness and Performance: A Unitary SART Index Predicts Performance on a Simulated ATC Task", in Human Factors in Aviation Operations. Fuller R., Johnston N., and McDonald N. (Eds) "Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Vol. 3. 1994, Aldershot, Avebury Aviation.

13. Nygren, T.E. ' Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload', Human Factors, 33 (1), 17-33. 1991.

14. Taylor, R.M. and Selcon S.J., "Operator and Automation Capability Analysis: Picking the Right Team". AGARD CP 520, April 1993, Paper 20.

15. Muir, B.M. "Trust Between Humans and Machines, and the Design of Decision Aids. Int. J. Man-Machine Studies, *27*, 1987, pp 527-539.

16. Lerch, F. and Prietula, M. " How Do We Trust Machine Advice?" In "Designing and Using Human-Computer Interfaces and Knowledge-Based Systems". Salvendy, G. and Smith, M. Eds. 1989, Amsterdam, Elsevier.

17. Taylor, R.M. "Trust and Awareness in Human-Electronic Crew Teamwork". In " The Human-Electronic Crew: Can They Work Together?" WRDC-TR-89-7008, Wright Patterson AFB, OH., 1988.

18. Riley, V. "Modelling the Dynamics of Pilot Interaction with an Electronic Crew". In " The Human-Electronic Crew: Is the Team Maturing?" WL-TR-92-3078, Wright Patterson AFB, OH, July 1992, pp 103-107.

19. Riley, V. "A Theory of Operator Reliance On Automation". In "Human Performance in Automated Systems: Current Research and Trends", Mouloua, M. and Parasuraman, R. (Eds), Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1994.

20. Lee, J.D. and Moray, N. "Trust, Control Strategies and Allocation of Function in Human Machine Systems". Ergonomics, *35*, pp 1243-1270.

21. Carmody, M.A. and Gluckman, J.P. "Task Specific Effects of Automation and Automation Failure on Performance, Workload and Situational Awareness", Proc.of 7th Int'l Symp. on Av. Psych., Ohio State Univ., 1993, Vol 1, pp 167-171.

22. Parasuraman, R., Mouloua, M., Molloy, R., and Hilburn, B. "Adaptive Function Allocation Reduces Costs of Static Automation", In "Proc. of 7th Intern'l Symp. on Av. Psych., Ohio State Univ., 1993, Vol 1, pp 178-181.

23. Rouse, W.B. "Twenty Years of Adaptive Aiding: Origins of the Concept and Lessons Learned". In "Human Performance in Automated Systems: Current Research and Trends", Mouloua, M. and Parasuraman, R. (Eds), Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1994, pp 28-32.33.

24. Trust and Adaptation Failure: An Experimental Study of Unco-operation awareness". *in*, "The Human Electronic Crew: Can We Trust the Team?" Proceedings of the 3rd International Workshop on Human Computer Teamwork. DRA CHS Report DRA/CHS/HS3/TR95001/02, January 1995, 93-98.

25. Comstock, J.R. and Arnegard, R.J. "Multi-Attribute Task Battery. Preliminary Technical Report", NASA Langley Research Centre, Hampton, VA. 1992.

26. Sen, T. and Boe, W.J. "Confidence and Accuracy in Judgements Using Computer Displayed Information". Beh. and Info. Tech., *10*, 1, 1991, pp 53-64.

27. Proctor, R.W. A forty year perspective on compatibility phenomena: A panel in honour of Arnold M Small. Proceedings of the HFS 34th Annual Meeting, 1990: 1444-1446.

28. Fitts P.M. and Seeger C.M. S-R compatibility: Spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 1953, 46, 199-210.

29. McCormick, E.J. and Sanders, M.S. Human Factors In Engineering And Design, 1982, New York: McGraw-Hill.

30. Wickens C.D. Engineering Psychology and Human Performance, 1984, Columbus: Merill.

31. Andre AD and Wickens CD. Compatibility and consistency in display-control systems: Implications for aircraft decision aid design. Human Factors, 1992, *34*, 6, pp 639-653.

32. Noble, D. Schema-based knowledge elicitation for planning and situation assessment aids. IEEE Transactions on Systems, Man, and Cybernetics. 1989. *19*, pp. 473-482.

33. Dix A., Finlay J., Abowd G., and Beale R. "Human Computer Interaction". New York: Prentice Hall, 1993.

34. Chechile RA. A review of ANNETS; A model of cognitive complexity of displays. Proceedings of HFS 36th Annual Meeting 1992, pp 1176-1180.

35. Chechile RA, Egglestone RG, Fleischman RN and Sasseville AM. Modelling the cognitive content of Displays. Human Factors. 1989, *31*, 1, pp. 31-43.

SITUATIONAL AWARENESS MEASUREMENT IN COCKPIT EVALUATION TRIALS

Barry McGuinness
Human Factors Department
Sowerby Research Centre
British Aerospace
FPC 267, PO Box 5, Filton
Bristol BS12 7QW
UK.

SUMMARY

Maintaining good situational awareness has always been a necessary task for pilots. However, the influx of more and more new technological capabilities in aircraft systems and the growing demands made by these on the pilot's attention has highlighted the vulnerability of situational awareness and its critical effects on pilot performance and mission success. In developing and evaluating novel displays and other new cockpit systems, measuring any differences in the pilot's situational awareness as well as performance has become a growing priority. We therefore require suitable tools for effective situational awareness measurement in simulation research rigs. This paper presents several insights into measuring situational awareness in cockpit systems research, and dicusses in detail the rationale for a proposed test-battery approach that incorporates a variety of measurement techniques.

1. INTRODUCTION

The fighter pilot must aim to maintain optimum situational awareness (SA) throughout a mission, even under stressful, high-workload conditions. Time spent acquiring SA must compete for attentional resources involved in making and executing effective decisions. Ideally, then, the information conveyed by the pilot's visual and auditory display systems should help him maintain good SA without the risk of excessive workload. Yet technological advances and more intensive task demands continue to remove pilots from direct sensory contact with the environment.

Increasingly, the pilot has to integrate information about his situation from multiple instruments, sensors and computer-generated images. The pilot's SA can therefore be helped or hindered by the design of specific cockpit systems. If we can ensure good SA through systems design, we should also be able to enhance performance effectiveness and decrease the likelihood of costly and disastrous mishaps. To gain a better understanding of how SA is affected by cockpit systems, and to be able to test any SA-enhancing methods objectively, we have a need for tools which can provide accurate, valid and reliable measurement of SA.

In the Sowerby Research Centre we are frequently called upon to provide human factors expertise in the design, running and analysis of advanced cockpit systems evaluation trials that make use of sophisticated simulation rigs. Because situational awareness is an aspect of growing concern, we have begun to explore methods of SA measurement that are suitable for trials of this nature. Similar work is, of course, underway at other centres — a good example is Northrop's SAGAT tool (Ref. 1), but this built-in query system is not accessible to us.

Through studies of the literature and some preparatory work, we have come to take a closer look at what situational awareness actually is and how it relates to a pilot's task performance. In the process, we have gained theoretical and practical insights into the design and application of SA measurement techniques to cockpit systems research in simulation trials. These are the topic of this paper.

2. A CLOSER LOOK AT SITUATIONAL AWARENESS

It is widely recognized that SA (like workload) is a multi-faceted aspect of a pilot's capabilities. One fundamental distinction to be made is between the *contents* of SA (i.e. the actual knowledge product that has been assembled) and the *processes* that go into acquiring and using SA (such as perception, interpretation, prediction, decision-making and so on). It is useful to make this content-process distinction explicit because some studies have been more concerned with SA knowledge contents and others with SA-related processes (see Table 1).

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

SA knowledge contents	*SA-related processes*
a. **Spatial** "Where everything is" — locations, distances, heights, relative to self (egocentric) and/or relative to world (exocentric).	a. **Sensing** Actively monitoring the appropriate sources of sensory data (visual, auditory, and so on) in order to observe developments and detect relevant changes. If anything novel, unexpected, ambiguous or significant is detected, selective attention is now required for further processing.
b. **Temporal** "What's happening when" — long-term planned schedule, near-future expected/projected events, trends, arrival times, deadlines, windows of opportunity, time pressures, possible changes, revisions to schedule.	b. **Identification** The new information is identified in terms of the pilot's knowledge, experience and expectations (e.g., based on his current SA).
c. **States & events** "What's new and what's what" — external: sensor detections, tactical developments, threat status; internal: systems status (fuel, power, weapons), alerts, warnings, emergencies.	c. **Interpretation** Determining the probable meanings of the perceived information in terms of current goals and other known factors.
d. **Identities** "Who everybody is" — friendlies, flight, support, enemies, targets, threats, neutrals, unknowns.	d. **Inference** Using his knowledge and experience, the pilot can try to infer the likely states, intentions and knowledge of others.
e. **Behaviours** "What everybody's doing right now" — manoeuvres, actions, unfolding patterns, approaching, threatening, attacking, evading.	e. **Integration** The interpreted information is itself incorporated into the pilot's SA — his mental model or 'picture'.
f. **Meanings** "What it all means" — interpreted probabilities, threat levels, risk levels, safety margins, significance of data, links between different factors, the "big picture".	f. **Projection** Making predictions of likely outcomes and events based on current knowledge and past experience.
g. **Intentions** "What everybody's aiming to do" — friendly intentions, enemy intentions.	g. **Metacognition** The capacity to be aware of one's own knowledge; the pilot can identify gaps in his own (or others') SA and assess his SA 'confidence level'.
h. **Response options** "What my choices are" — possible courses of action available.	h. **Analysis** Analyzing the specific *response options* (manoeuvres, tactics, courses of action, and so on) that are available, evaluating pros and cons.
i. **Projected situations** "What might happen" — likely outcomes of own and others' actions.	i. **Decision-making** Selecting the appropriate response from those available with reference to goals and priorities.
j. **Metacognition** "Knowing what's known" — awareness of quantity/quality of own knowledge state; subjective confidence in own SA; awareness of *unknown* aspects, estimating what others know and don't know.	
k. **Goals** "What I'm now aiming to do" — selected response defines new objectives, intentions, priorities.	

TABLE 1: A listing of SA knowledge contents and SA-related processes

2.1 SA contents

The situation of which the pilot needs to be aware comprises all factors that are important to current mission objectives and task requirements — states, events, locations, changes, trends and so on. A fuller compilation of SA contents, based on current thinking in the area (both published and in-house), is given in the first column of Table 1.

The required data come to the pilot from several sources: from the environment or outside world, perceived directly or via sensor displays; from crewmembers, ground operators and other participants via communications links (including digital datalink); and from the aircraft systems, mostly presented via cockpit displays and instruments. There may of course be some very relevant data that are physically undetectable to the pilot — threats not yet picked up on radar, for instance.

While it is the pilot's job to acquire and process all available data, systems designers should aim to ensure that the right data are made available to the pilot at the right time so that the pilot can achieve optimal SA.

2.2 SA processes

The psychological mechanisms involved in establishing and maintaining SA logically involve a hierarchy of cognitive processes which can be modelled in various ways. At the lowest level are elementary and automatic sensory processes, such as motion detection. Higher up are processes involved in perceptual integration, putting together the overall pattern. At the highest level are the most active and conscious processes: evaluating the perceived pattern with respect to current goals and making predictions. At this level, the SA content contributes directly to conscious decision-making.

These processes can be generally thought of not so much as a sequence of discrete steps but as a constant flow of information through a hierarchy of levels of analysis, transforming the raw sensory data into integrated knowledge. However, it is easiest to describe the processes sequentially. The second column in Table 1 attempts to give a comprehensive inventory of SA-related processes as they are currently conceived in the general literature and by ourselves.

By examining the pilot's SA-related processes together with the knowledge that these act on and produce, we can posit that the interactions of the two, contents and processes, progressively give rise to several key components of SA (as illustrated in Fig 1), namely:

1. **Perception** of the explicit situation
2. **Comprehension** of the 'big picture'
3. **Projection** of future situations
4. **Metacognition** i.e. self-assessment of own SA
5. **Response Selection** choice of performance goals

Some of these components deliberately overlap with Endsley's (Ref. 1) description of three levels of SA:

> **level 1** — *perception* of the physical elements of the situation (who, what, where, etc.);
> **level 2** — *comprehension* of the significance of objects and events within the situation (intentions, threat levels, options);
> **level 3** — *projection* of future events based on the current situation.

In addition to these three, however, we can include another two components of SA that seem relevant — *metacognition* and *response selection* — as listed above. If we are to develop a tool for comprehensive SA measurement, then all five components need to be addressed.

3. A VARIETY OF MEASUREMENT TECHNIQUES

The challenge of measuring SA in cockpit evaluation trials is to determine how much of the available task-critical information the pilot is (or has been) able to perceive, comprehend and use appropriately. In other words, to what extent is the pilot's mental model a reflection of his real situation? This includes not only the explicit aspects of the situation as it is physically perceived, but also the implicit aspects which come from the pilot's own expert judgement, such as inferring the probable intentions of others from the context of their behaviours.

Only one approach so far, *knowledge elicitation*, seems to offer direct measurement of content. The SAGAT tool is a well-known example of this type of method, aimed at addressing the perception, comprehension and projection components. At Sowerby we have begun to explore ways of optimising knowledge-elicitation techniques for use in our own evaluation trials, together with methods for evaluating metacognition and response selection. The following summarises current insights.

3.1 Knowledge elicitation

The typical method of knowledge elicitation is to give the pilot a sequence of factual questions (probes, queries) which, in theory, he should be able to answer correctly. The fewer questions he is able to answer correctly, the lower is his SA score.

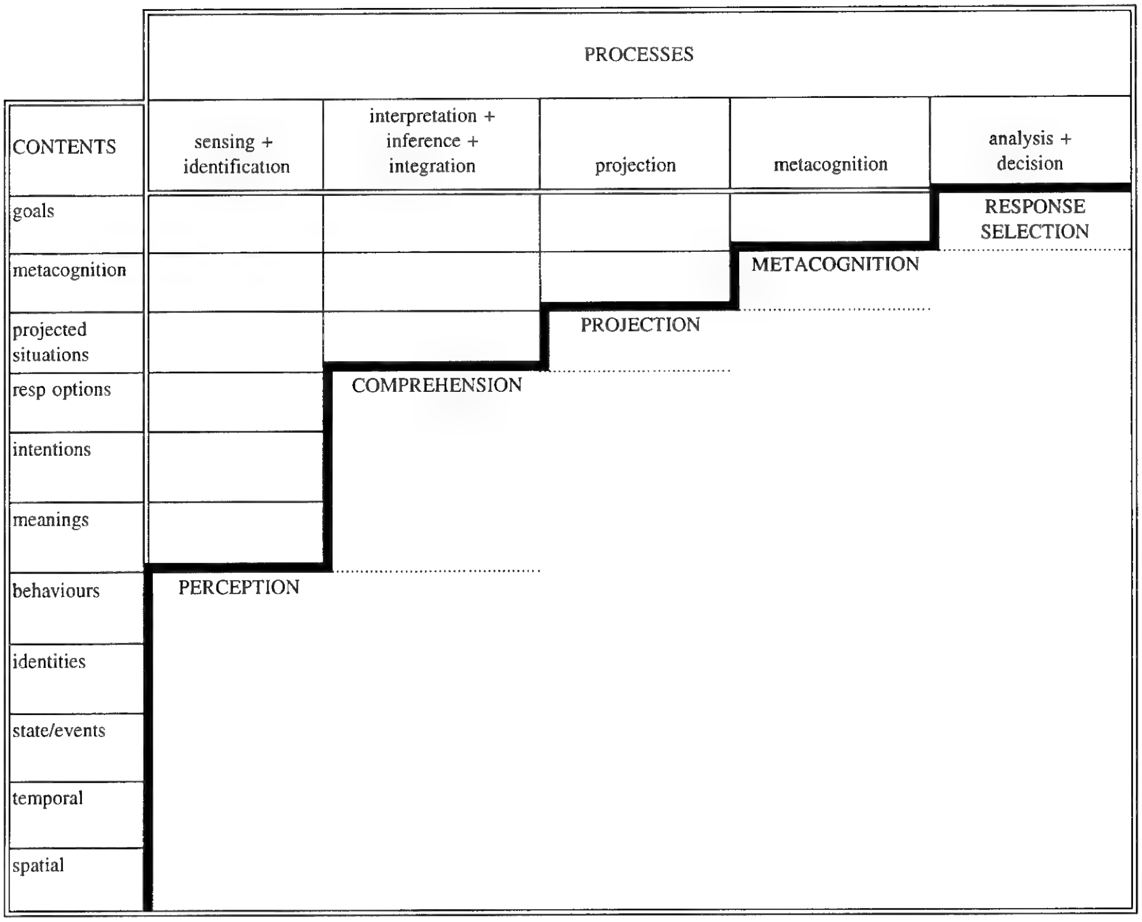


FIGURE 1: Hypothetical progressive build-up of SA components

The probes can be presented either 'live' during the actual trial (mid-run) or as soon as the trial is completed (post-run). Mid-run probes involve a disruption of the pilot's tasks, but have the advantage that they involve immediate recall and can be asked repeatedly, effectively providing a 'profile' of SA accuracy over time. Post-run probes can only provide a global score for the entire trial, but they have the advantage of being completely non-intrusive.

Questions can be multiple-choice or they can be asked without any choices being offered, depending upon requirements. The advantage of multiple-choice questions is that they confine the answers to a finite range — the responses are either right or wrong. The advantage of more direct questions is that they elicit precise figures which can be evaluated against known data.

3.1.1 Verbal mid-run probes
A simple and non-interruptive method of mid-run probing is to provide spoken queries via the pilot's communication system, these sounding as if they could be coming from a fellow crew member, wingman or air traffic controller (Ref. 3). It is necessary, however, to have the capability of 'hiding' the specific data being queried out of the pilot's sight, essentially by blanking off the relevant display alphanumerics and symbols for the duration of the probe. For example, if asking: "What is your current heading?", any heading indicators on the head-up and head-down displays must be temporarily removed or concealed. Probes need to be fairly well spaced apart, as intensive questioning may contribute artificially to the pilot's workload. The questions administered in this way need to be straight rather than multiple-choice, so this technique is particularly suited to eliciting the pilot's current perceptions of the explicit data. Responses can then be compared with actual data stored at the time of questioning.

3.1.2 Interruptive mid-run probes
The mid-run technique used by SAGAT involves temporarily freezing and shutting off the entire simulation while multiple-choice questions are presented on a touch-screen. Research by Endsley

(Ref. 2) suggests that freezing the simulation in this way has no apparent impact on performance, and a freeze-delay of up to 6 minutes has no apparent effect on pilots' ability to report their SA knowledge. Subjects' multiple-choice responses to SAGAT questions — such as "AIRCRAFT A1 IS: OFFENSIVE / DEFENSIVE / NEUTRAL" — are immediately stored by computer and are subsequently converted into an SA score. This method is particularly suited to testing the pilot's comprehension of the situation and can also incorporate probes of SA projection.

3.1.3 Spatial tests

Because spatial knowledge plays such a key role in SA, a complementary method to factual queries is to test the pilot's spatial awareness during a simulation freeze by having him indicate spatial layout and positions, e.g. on a map. At Sowerby, Fuchs (Ref. 4) has proposed a method he calls "*recreating the big picture*", by which a pilot could use a model to reconstruct his own position and those of other participants. The model could be physical (e.g., positioning a model aircraft within a 3D environment) or it could be an interactive computer-generated image. Comparing recreated positions with actual coordinates (stored on computer) will provide a measure of spatial accuracy.

3.1.4 Post-run probes

An alternative to mid-run probing is to elicit SA knowledge post-flight. The problem here is that the pilot's memory is involved, rather than SA per se. Information that no longers needs to be known may well be forgotten. Endsley (Ref. 2) suggests that level 1 information (perception) is briefly held in working memory just long enough for deeper levels of meaning to be extracted from it; therefore it is easily vulnerable to being forgotten. Levels 2 and 3 (comprehension and projection), on the other hand, involve a deeper level of processing; the knowledge is more effectively laid down in long-term memory. So while solutions to levels 2 and 3 queries might be readily retrieved with delayed recall, answers to level 1 queries are probably only available through immediate recall. In other words, items concerning *perception* need to be probed mid-run, while items concerning *comprehension* and *projection* could be probed post-run if feasible. The problem then is that the pilot now has knowledge of the outcome of the flight, so projection probes in particular would need to be very carefully designed.

3.1.5 Selection of probes

The most crucial challenge with knowledge probing in general is that *the probes used must be relevant to the pilot's SA*. A question is useless if it addresses something that the pilot does not consciously include as an aspect of his SA at that time. Arriving at an adequate mental model of the situation involves acquiring the relevant information but then integrating it into current SA in accordance with mission/task goals. At any point within any mission, there are many system and environmental variables that are relevant to the pilot's task and that can influence what he may choose to do next. Part of the pilot's task is to monitor these variables through the available channels and sources of information, and to integrate all the relevant data so as to formulate the most effective decisions. As Metalis (Ref. 5) has noted:

> A pilot who has SA is like an "expert" who can look at a huge array of discrete stimuli and immediately integrate them into "chunks", or meaningful bytes of knowledge upon which he can base appropriate action.

This 'chunking' appears to be a key attribute of SA knowledge beyond the level of initial perception.

In the cockpit, most of the key variables will be represented as continuous data (e.g., bearing 137, fuel 55%, altitude 2300 feet, etc.). As far as the pilot is concerned, however, each relevant variable reveals a particular *state* (e.g., HIGH/LOW/OK). The pilot perceives the numerical data on a display (e.g., altitude 2300 feet), but rapidly translates the figures into a meaningful, qualitative state (e.g., "TOO HIGH").

A problem with querying the pilot's awareness of the detailed aspects of the situation, then, is that the pilot is not really concerned with, say, his altitude reading as such but with whether his altitude is going too low, too high or is at about the right level. A very specific post-run probe like "what was your closing speed at the time of missile launch?" may be deemed by pilots to have little meaning in terms of SA; the pilot's real concern is more likely to be the fact that he got into the right position for a firing opportunity — the numerical details will no longer matter.

3.1.6 Knowledge integration

Beyond the initial stage of perception, then, the raw data presented by cockpit displays are not retained as part of SA; only their *meanings* are integrated into the mental model. In other words, the pilot's SA with respect to the comprehension of numerical variables involves meaningful 'fuzzy concepts' rather than precise figures. It appears, therefore, that it is these semantic states, combined with analogic visuospatial representations, that the pilot uses in order to make decisions. These states are not arbitrary simplifications but are meaningful chunks,

directly relevant to the current tasks. For example, one reason for monitoring altitude is to detect the state "TOO LOW — IN DANGER OF HITTING THE GROUND." Even spatial and temporal knowledge ultimately feed into a perceived 'state': either the current arrival time is OK or it is not; the rate of descent is too high, too low or OK; the location of a potential threat is safe or dangerous; and so on.

In the end, SA is about understanding whether one's present course of action is acceptable in terms of one's goals, or needs to be changed. Knowing this allows a pilot to make effective decisions. It is therefore important to consult the expert knowledge of pilots themselves in order to determine the most appopriate SA questions for a given trial. If the questions are to be asked mid-run, then the *times* at which those items of information are actually relevant to the pilot should also be determined.

There is a clear need for more research into the effective use of knowledge-elicitation probes (Ref. 6). This is important because knowledge-elicitation is probably the most direct SA measure. With proper development, it can not only indicate exactly how accurately the pilot perceives and understands the situation, but it can also serve as a validation baseline in studies of more indirect measures.

3.2 Subjective ratings

In addition to knowledge probing, a less direct approach to SA measurement is to measure something else that correlates with some aspect of SA and then infer the pilot's SA from that (e.g, physiological measurement). *Subjective ratings* techniques typically obtain an estimate of the pilot's SA that is derived from his own subjective sense. This provides, in effect, a semi-direct measure of the 'metacognition' component of SA.

Subjective ratings techniques are an important complement to measures of SA content provided by knowledge elicitation, as a pilot may do rather poorly on the knowledge queries yet still *feel* confident that his SA is adequate for the tasks at hand. This could either be because the selected knowledge queries are inappropriate, or because the pilot actually is oblivious to some important information.

3.2.1 SART

Taylor and Selcon's SART approach has become a popular subjective-ratings tool (Refs. 7, 8). This can be administered either as a 10-scale post-run version or as a simpler 3-scale mid-run version. The tool is often found to be less than perfect, however. In particular, the terminology used to define the rating scales is psychological rather than 'plain English'. Pilots must be given a thorough briefing to arrive at some understanding of what they are being asked to rate, but even this still leaves room for misreadings and misinterpretations. Another minor difficulty with SART is its broad scope — most of the scales it includes refer to aspects of workload rather than to situational awareness per se.

3.2.2 A simple rating scale

What is required at SRC for sampling subjective SA (metacognition) in cockpit evaluation trials is a simple, rapid-response rating scale that is very easily understood by the users. There is no reason why we should not be able to address the pilots' own intuitive understanding of SA, which might be described as: "KNOWING WHAT'S GOING ON SO YOU CAN FIGURE OUT WHAT TO DO!" (Ref. 9). Pilots could then simply be asked mid-run to rate on a scale of, say, 0-10, the extent to which they feel confident that they know what's going on (so that they can figure out what to do). Such a scale could be responded to throughout a trial at regular intervals, providing at the end of the run a profile of how the subject's subjective sense of SA varied during the flight. These regular, periodic ratings could be administered either verbally or via some computer-controlled device in the cockpit.

3.3 Response selection

SA is crucially related to decision-making — the whole point of having good SA is determining the best thing to do next. A pilot is expected to be able to detect and interpret all relevant information and use it to make effective decisions: anticipating likely developments, knowing all the available options and selecting the best one. Yet decisions are only as good as the pilot's situational awareness, so any inaccuracies in perception can lead to poor choices of action. While the situation itself determines which response options are realistic, it is the pilot's own SA (based on his subjective perceptions of the situation) that determines which options he *believes* to be feasible and which not. If his SA is accurate, the pilot should be able to select the best responses.

3.3.1 Verbal protocol analysis

It is difficult to see how response selections can be evaluated quantitatively during full-blown simulation trials. Analysing a pilot's decisions qualitatively, however, should shed some light on this important aspect of SA. One way to achieve this is by post-run analysis of the pilot's performance, examining the choices and actions that have been made. A crucial aid to this process would be the use of *verbal protocol analysis* — that is, having the subject 'think aloud' during the trial whenever he considers his response options and makes decisions. If the subject also collaborates in the post-run analysis,

perhaps with the aid of a computer-graphic 'replay' of the flight run, this will help reveal *why* particular decisions were made at particular times. This should provide useful insight into each individual pilot's acquisition and use of SA.

3.4 Query response time

One additional method that has not received much attention is to record the pilot's *query response time* when applying knowledge-elicitation or subjective-ratings techniques. In theory, this will give an indication of the extent to which the SA-related processing needed to give an answer has already been completed. For example, the pilot might be giving a correct response to a comprehension probe, but was the information already known to him (as part of his current SA) or has he had to work it out? If the answer to a probe is readily known, then the response time should be minimal; if the answer must be constructed from available information, then the time to complete this process will, in theory, reflect the state of SA at that time.

With regard to the major components of SA, three of these—perception, comprehension and projection—can be addressed by knowledge-elicitation methods, while a fourth—metacognition—is measurable using subjective-ratings techniques. In all these cases, we can also take the query response time as a measure of SA processing completion.

4. SITUATIONAL AWARENESS MEASUREMENT TOOL: SRC APPROACH

Ideally, we at SRC would like a tool for measuring SA that is portable, easy-to-use, valid, reliable, sensitive, and has demonstrable predictive power. Current tools intended to provide SA measurement fall well short of this ideal. Part of the problem is that SA is a multi-faceted psychological phenomenon, so a variety of different but coordinated measurement approaches may be required in order to get the full picture.

4.1 A battery of metrics

Using direct knowledge probes, we can sample the pilot's awareness and understanding of not only what *is* happening (perception, comprehension) but also what is *going* to happen (projection). Verbal protocol analysis can reveal what the pilot thinks should be *done* about it (response selection), while subjective ratings can give an indication of how much SA the pilot himself believes he really has (metacognition).

The optimum configuration of measurement techniques will probably be influenced by the specific demands of the trial objectives and the nature of the given simulation systems. A typical configuration, however, may to be administer non-interruptive mid-run verbal probes for general data perception, with interruptive mid-run multiple-choice queries under computer control for comprehension (including spatial awareness) and projection, plus non-interruptive mid-run subjective ratings at regular intervals for metacognition, and in all cases recording the time it takes the subject to formulate a response. In addition, subjects can be instructed in verbalising their thought processes aloud throughout the flight, thus providing a means to study response selections.

The more these techniques can be harmonised with the existing simulation facilities and procedures, the better. At the Sowerby Research Centre we are embarking on the development of a battery of SA metrics that can be readily used in BAe simulation trials (see Table 2 for a summary). The advantage of such a tool would be that it provides a consistent and structured measure of SA over multiple trials, allowing for comparative system evaluations. The measures can be integrated with other metrics and measurement systems which are in place, such as performance recording and workload measurement.

4.2 Conclusions

It may never be possible to isolate a single absolute measure of SA. Insights and strategies such as those presented here, however, should enable us to proceed with constructing a multi-factorial approach that will provide results of sufficient breadth and objectivity for the purposes of cockpit systems evaluation in simulator-based research. Techniques like this will help us to more objectively evaluate cockpit display systems in terms of their impact on pilot situational awareness.

SA component	Measurement approach	Proposed methods
PERCEPTION	knowledge probes	non-interruptive mid-run probes straight questions: — spatial, temporal, states/events, identities, behaviours verbal administration
COMPREHENSION	knowledge probes	interruptive mid-run probes factual multiple choice questions: — computerised administration (SAGAT-like) spatial test: — recreation of the big picture (paper/model/computer?) possibly post-run probes
PROJECTION	knowledge probes	interruptive mid-run probes as above, but focusing on future rather than current situation
METACOGNITION	subjective ratings	non-interruptive mid-run probes periodic self-rating of SA level (verbal/computer?) possibly post-run rating (SART-like)
RESPONSE SELECTION	verbal protocol analysis	post-run analysis of pilot's mid-run verbalisations preferably with computer-generated replay

TABLE 2: Proposed battery of SA measurement techniques

REFERENCES

1. Endlsey, M.R. "A methodology for the objective measurement of situation awareness", *in* "Situational Awareness in Aerospace Operations", AGARD CP 478, 1989, Paper 1.

2. Endsley, M.R. "Predictive utility of an objective measure of situation awareness", *in* the Proceedings of the Human Factors Society 34th Annual Meeting, 1990, 41-45.

3. Sarter, N.B. and Woods, D.L., "Situation awareness: A critical but ill-defined phenomenon", International Journal of Aviation Psychology, *1*, 1, 1991, 45-57.

4. Fuchs, M., "The measurement of situational awareness", JS11816, BAe Sowerby Research Centre, 1993.

5. Metalis, S.A., "Assessment of pilot situational awareness: measurement via simulation", *in* the Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, 113-117.

6. Vidulich, M.A., Stratton, M., Crabtree, M. and Wilson, G., "Pereformance-based and physiological measures of situational awareness", Aviation, Space, and Environmental Medicine, May 1994, A7-A12.

7. Taylor, R.M., "Situation Awareness Rating Technique (SART): the development of a tool for aircrew systems design", *in* "Situational Awareness in Aerospace Operations", AGARD CP 478, 1990, Paper 3.

8. Selcon, S.J. and Taylor, R.M., "Evaluation of the Situational Awareness Rating Technique (SART) as a tool for aircrew systems design", *in* "Situational Awareness in Aerospace Operations", AGARD CP 478, 1990, Paper 5.

9. Adam, E.C., "Fighter cockpits of the future", *in* "12th DASC", the Proceedings of the 1993 IEEE/AIAA Digital Avionics Systems Conference, 318-323.

Development of Criterion Measures of Situation Awareness for Use in Operational Fighter Squadrons

Wayne L. Waag
Aircrew Training Research Division
Armstrong Laboratory
6001 S. Power Rd.
Mesa, AZ 85206-0904
USA

Michael R. Houck
University of Dayton Research Institute
PO Box 2020
Higley, AZ 85236
USA

SUMMARY

This paper describes the development of three Situation Awareness Rating Scales (SARS) that were used to measure pilot performance in an operational fighter environment. These instruments rated situation awareness (SA) from three perspectives: supervisors, peers, and self-report. SARS data were gathered from 239 mission-ready USAF F-15C pilots from 11 operational squadrons. Reliabilities of the SARS were quite high as measured by their internal consistency (.97 to .99) and inter-rater agreement (.84). Correlations between the supervisory and peer SARS were strongly positive (.85 to .87), while correlations with the self-report SARS were positive, but smaller (.50 to .58). A composite SA score was developed from the supervisory and peer SARS using a principal components analysis. The resulting score was found to be highly related to previous flight experience and current flight qualification. A prediction equation derived from available background and experience factors accounted for 73% of its variance. Implications for use of the composite SA score as a criterion measure are discussed.

1 INTRODUCTION

Study Background. In 1991, the US Air Force Chief of Staff posed a series of questions concerning SA that led to the present investigation. First of all, what is SA? Can it be objectively measured? Is SA learned or does it represent a basic ability or characteristic that some pilots have and others do not? From a research standpoint, these questions translate into issues of measurement, selection, and training. The Armstrong Laboratory was subsequently tasked with providing research answers to these questions. A research investigation was initiated that had three goals: first, to develop and validate tools for reliably measuring SA; second, to identify basic cognitive and psychomotor abilities that are associated with pilots judged to have good SA; and third, to determine if SA can be learned, and if so, to identify areas where cost-effective training tools might be developed and employed. An overview of the investigation can be found in this report in the paper by McMillan, Bushman, and Judge (1).

The general approach was to first develop criterion measures of SA based upon performance ratings collected within an operational flying environment. These measures were necessary for two reasons. First, they would serve as criterion measures against which to validate a battery of basic ability tests considered relevant to SA, thereby addressing the question of basic human abilities. The results of this part of the study can also be found in this report in the paper by Carretta and Ree (2). Second, these measures would serve as a means of selecting a sample of pilots who would participate in a simulation phase of the effort. During that phase, simulated air combat mission scenarios were developed for assessing SA and objective measures of performance gathered in an attempt to determine those characteristics that distinguish pilots with good SA. These data would be used to identify areas where training tools might be developed. The results of this part of the study can be also be found in this report in the paper by Waag, Houck, Greschke, and Raspotnik (3). This paper presents the results of only the first phase of the program, namely, the development of tools for measuring SA within an operational fighter environment.

The approach to developing measurement tools was largely dictated by the definition of SA adopted at the outset of the study, the intended use of the data, and practical constraints involved in gathering data on mission-ready aircrew. In response to the question, "what is it?", the Air Staff produced the following operator's definition of SA: "a pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception. (4)" While other definitions of SA within the literature focus primarily on processes underlying the assessment of the situation (5), our working definition also included forecasting, decision making, and task execution. As such, it was viewed as a fairly global operational concept that encompasses much of the domain of air combat proficiency. Since the data were to be used primarily as a criterion against which to determine relationships with basic ability measures, fairly large numbers of subjects would be required. This requirement further restricted the types of measures to those that could be gathered in a fairly quick, non-invasive manner since available pilot time within any operational flying environment is quite limited. While a number of measurement tools had been developed to measure SA within a highly controlled, simulated flight environment (6,7), these could not be used due to practical constraints. Consequently, previous efforts to develop criterion measures of combat effectiveness were considered (8).

Measurement of Combat Effectiveness. Attempts to measure and predict combat effectiveness have a long history dating

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

back to the second world war. The interested reader is referred to Youngling *et al.* (8) who conducted an extensive review of this literature. In essence, there are two problems that must be addressed: first, the definition of the criterion for combat effectiveness, and second, the search for measures that are predictive of that criterion. In general, four types of criteria have been used: objective outcome measures such as kills, bombing scores, etc.; direct and systematic observations of mission performance; administrative actions such as failure to complete a fighter tour; and qualitative ratings of overall ability. On the predictor side, a variety of potential indicators of combat effectiveness have been explored; including basic aptitude, biographical factors including flight experience, a variety of personality and motivational factors, perceptual-motor abilities, and knowledge and skills directly related to aviation.

In general, only very modest relationships have been obtained. Of the predictor sets that have been evaluated, those measures related to previous flight experience seem to be most consistently related to combat effectiveness as measured by combat kills. Strawbridge and Kahn (9) and Torrance *et al.* (10) reported correlations with previous flight experience in the range of .30 to .40. Correlations with aptitude test scores and other perceptual-motor tests were substantially lower, with most failing to reach statistical significance. DeLeon (11) summarized the results of the Red Baron studies that were conducted during the Vietnam conflict. Flight experience in terms of total flight hours, total fighter hours, and hours in the combat aircraft was found related to combat success, although the degree of these relationships was fairly small. DeLeon (11) concludes that "at best, experience appears to be only a vague measure of a pilot's air-to-air combat skills (pg 16)".

In summary, previous studies have reported the highest relationships between flight experience and criteria involving actual combat success, i.e., kills. However, such criteria were not available for the present study. Nor was it feasible for reasons of time, cost, and lack of combat realism to gather data based on actual mission performance in the aircraft under the highly controlled conditions of an instrumented range environment as suggested by Youngling *et al.* (8). For practical reasons, the only alternative was to develop criterion measures based upon human judgments. Unfortunately, the use of subjective ratings of overall ability as the criterion of combat effectiveness has produced few statistically significant relationships with predictor sets that have been used to date. For example, Lepley (12) found only one significant correlation for his test battery with subjective ratings of ability. Shannon and Waag (13,14) met with limited success in an attempt to relate background and experience factors to operational performance. In this case, squadron commander ratings of mission-critical performance dimensions were the criterion measures. Results indicated that flight experience was the best predictor of criterion performance. Undergraduate Pilot Training grades for Formation and Tactics were also found related to such ratings. However, the overall magnitude of the relationship was fairly low with a multiple correlation of all background and experience factors reaching only .35.

<u>Measurement Approach</u>. In general, three types of performance ratings have been used. The most common, and also most cited in the literature, has been supervisory ratings (15). The two other types include peer ratings and self ratings. In fact, the use of peer ratings for combat aviation dates back to World War II when Jenkins *et al.* (16) developed a "combat" criterion for the US Navy based largely upon peer nominations. In an extensive literature review, Landy and Farr (15) conclude that previous research studies have not found very high correlations among these three types of ratings. Moreover, it is difficult to select one approach as best since the literature does not suggest any of these to be more valid than the others. For these reasons, it was decided to develop three SA Rating Scales (SARS) and gather supervisory, peer, and self-report data. Moreover, it was decided to use a simple graphic scale since the literature is equivocal regarding more elaborate procedures such as behaviorally anchored rating scales (15).

What seemed most critical, however, were the actual dimensions that were to be rated and the development of clear definitions for each. To characterize the domain of air combat, it was necessary first to identify and describe the critical activities required of the fighter pilot to maintain good SA and complete his mission successfully. To this end, Houck *et al.* (17) conducted a cognitive task analysis of the attack portion of an F-15C air combat mission. This analysis relied primarily on the input of experienced fighter pilots and focused on critical air combat task categories that in previous research were rated by F-15C pilots as being most amenable to training in air combat simulations (18,19,20). The resulting analysis identified the significant types of decisions required of the flight members, the information required for making these decisions, and the observable activities the flight members performed to acquire this information. For the purposes of the present research, this classification provided a detailed description of optimum performance in air combat. The resulting classification was further analyzed by an experienced fighter pilot to derive those aspects of air combat operations judged most essential to SA. Paramount in this selection process was that the items must be observable in the context of day-to-day squadron training activities and subject to evaluation by other fighter pilots both in terms of their own performance and that of others. A further requirement was that the pilots must be able to assess these items in retrospect, based on performance observed to date. As a result of this analysis, 24 items organized in seven categories were produced. Categories included tactical game plan, system operation, communication, information interpretation, beyond-visual-range weapons employment, visual maneuvering, and general tactical employment. Because the 24 items were heavily weighted toward specific operational tasks, an additional 7 items were included to reflect more general traits which also were hypothesized to play a role in SA. These items were based on the study of fighter pilot combat effectiveness previously discussed (8). Concise definitions for

each item were developed with assistance from an experienced fighter pilot. The resulting list and definitions were reviewed and revised by several other experienced pilots to ensure accuracy and completeness. These 31 items and the 8 categories that they represent are presented in Table 1 and form the essence of the approach taken to the measurement of SA in the present study.

TABLE 1. ITEMS AND CATEGORIES USED IN SARS

1. GENERAL TRAITS
    - Discipline
    - Decisiveness
    - Tactical knowledge
    - Time-sharing ability
    - Spatial ability
    - Reasoning ability
2. TACTICAL GAME PLAN
    - Developing plan
    - Executing plan
    - Adjusting plan on-the-fly
3. SYSTEM OPERATION
    - Radar
    - Tactical electronic warfare system
    - Overall weapons system proficiency
4. COMMUNICATION
    - Quality (brevity, accuracy, timeliness)
    - Ability to effectively use information
5. INFORMATION INTERPRETATION
    - Interpreting vertical situation display
    - Interpreting threat warning system
    - Ability to use controller information
    - Integrating overall information
    - Radar sorting
    - Analyzing engagement geometry
    - Threat prioritization
6. TACTICAL EMPLOYMENT-BVR
    - Targeting decisions
    - Fire-point selection
7. TACTICAL EMPLOYMENT-VISUAL
    - Maintain track of bogeys/friendlies
    - Threat evaluation
    - Weapons employment
8. TACTICAL EMPLOYMENT-GENERAL
    - Assessing offensiveness/defensiveness
    - Lookout
    - Defensive reaction
    - Mutual support

Study Objectives. The purpose of the present investigation was to develop a set of tools for measuring SA within an operational fighter environment. Issues that are addressed in the present paper include: (1) reliability and validity of the SARS; (2) inter-relationships among the supervisory, peer, and self-report SARS; (3) development of a single composite SA score; and (4) external validity of the composite SA score as determined by relationships with previous flight experience factors.

2 METHOD

Subjects. The subjects were 239 mission-ready USAF F-15C pilots from 11 operational Fighter Squadrons. Mean, standard deviation, and range of flight hours were as follows: Total flight hours beyond undergraduate pilot training (1073, 590, 193 to 2805) and total flight hours in the F-15 (620, 257, 74 to 1400). Current flight qualifications of these pilots, in order of increasing experience and proficiency, included 62 wingmen, 67 two-ship leads, 43 four-ship leads, and 67 instructor pilots.

Materials. Five scales were developed for this study. The first was a questionnaire designed to obtain background and experience information such as flight hours, attendance at exercises, Desert Storm experience, etc. The second scale attempted to obtain information on the perceived importance of the 31 elements of SA. It required the subject to first give his own personal definition of SA and then to rate the importance of each of the 31 elements based on that definition. A simple six-point Likert scale was used.

The three remaining scales were developed to measure the SA ability of a pilot from three different perspectives: the self-report SARS, the peer SARS, and the supervisory SARS. Survey forms were custom designed and reproduced through an offset printing process to make use of computer-based data scanning technology. Each survey type was two pages. The first page contained printed instructions, scale description, and subject identification codes. The second page contained the actual rating scales.

For the self-report SARS, subjects rated their own ability on each of the 31 items in comparison with other F-15C pilots using a six-point scale. End-point anchors ranged from a low of "Acceptable", since all pilots were mission-ready, to a high of "Outstanding." The peer SARS required each subject to rate all other mission-ready pilots in his squadron. Each pilot listed on the peer SARS was rated on his general fighter pilot ability and SA ability using the same six-point scale. Once these ratings were completed, these pilots were then rank-ordered from highest to lowest in terms of their SA ability. A provision was included on the form for not rating a pilot if the rater felt he had insufficient knowledge of that particular individual. The supervisory SARS used the same 31 items and the six-point scale as the self-report SARS. Again, the reference was the relative ability of the ratee in comparison with other F-15C pilots.

The self-report and peer SARS were completed by all subjects within the sample. The supervisory SARS were completed by only a subset of subjects chosen to be raters, based upon their experience and supervisory positions. Raters within each squadron included: the Squadron Commander, Ops Officer, Assistant Ops Officer, Weapons Officer, and Stan-Eval Flight

Examiner (SEFE) who rated all mission-ready pilots within the Squadron; and the Flight Commanders, who only rated pilots within their flight as well as other Flight Commanders.

Procedures. The surveys were administered on location at each fighter squadron base. An elaborate numerical coding procedure was followed to ensure the confidentiality of each subject's data. The survey administrators briefed all subjects regarding the objectives of the research, scale description and item definitions, confidentiality procedures, and instructions for completing the surveys. Identification codes and dates on each survey were already filled in for each subject prior to administration. Each subject removed the surveys enclosed within the envelope, completed them, returned them to the envelope, and removed all name labels. These labels were given to the test administrator who destroyed them, thus leaving no name identifications within or outside the envelope.

Data regarding each subject's flight career experience were obtained directly from a computerized database and through responses to the background questionnaire administered to each subject. These data included flight hours and sorties by aircraft type, hours and sorties for both combat and combat support missions, current flight qualification, supervisory responsibilities, advanced fighter training, and participation in special fighter exercises and training simulations.

For the self-report SARS, 9 summary scores were produced. These included an overall score, which was the mean of all 31 items, and 8 category scores, which were the means of all items within a particular category. For the supervisory SARS, the same 9 summary scores were generated for each ratee as follows. First, the same 9 summary scores were computed for each rater's assessment of each ratee. Then, means for each summary score were computed across all raters and used as the final 9 supervisory SA scores for each ratee. For the peer SARS, 3 summary scores were produced for each ratee as follows. First, three scores were generated by each rater for each ratee, the ratings of fighter pilot ability and SA ability, and the rank order. Means of these three scores were then computed across all raters and used as the final peer SA scores for each ratee.

3 RESULTS

Pilot Validation of Rating Scale Approach. A concern at the beginning of the study was the question of whether the 31 elements used in the rating scales were in agreement with pilots internal views of SA. To partially answer this question, we compared the pilot's written definitions of SA with the results of the importance ratings wherein the subjects rated the importance of each of the 31 elements of SA. Of the 239 total pilots surveyed, 206 pilots provided a definition of SA. These definitions were transcribed into a computer text file and coded into several categories using the SHAPA (Version 2.0) protocol analysis software (21). This verbal analysis was accomplished through a two-stage process. The first stage involved an analysis of a randomly-selected subset of 25 of the pilot-produced SA definitions. Pilot definitions were decomposed into separate statements regarding SA and these statements were then analyzed and organized into categories based on common meaning. By way of this process, 12 categories of SA statements were produced. The second phase involved the analysis of all 206 pilot-produced definitions. Each pilot's SA definition was decomposed into separate statements regarding SA and then each statement was coded according to its representative category. The number of pilots making SA statements falling into each of the respective SA categories was tallied using SHAPA's data summary routines. Table 2 presents the seven most frequently cited statements from the written definitions.

TABLE 2. STATEMENTS FROM PILOT DEFINITIONS

Composite 3-D image of entire situation
Assimilation of information from multiple sources
Knowledge of spatial position or geometric relationships among tactical entities
Periodic mental update of dynamic situation
Prioritization of information and actions
Decision making quality and timeliness
Projection of situation in time

Table 3 presents the seven most important items from the 31 elements as determined by the mean of the assigned importance ratings. As shown, there is considerable agreement.

TABLE 3. MOST IMPORTANT ITEMS FROM SARS

Use of communication information
Information integration from multiple sources
Time-sharing ability
Maintaining track of bogies and friendlies
Adjusting plan on-the-fly
Spatial ability to mentally picture engagement
Lookout for threats from visual, RWR, VSD

SARS Reliability. The next set of analyses addressed the reliability of the SARS instruments. Two types of reliability were estimated, internal consistency and inter-rater agreement. First, internal consistency was estimated for the supervisory and self-report SARS by calculation of Cronbach's coefficient α. For the supervisory SARS, coefficient α was computed to be .99 for all 31 items. These results were based on the total number of supervisory SARS completed (N=1080). For the self-report SARS, α was computed to be .97 for all 31 items. Again, these were based on the total number of self-report SARS completed (N=235). Second, inter-rater agreement was estimated for the supervisory SARS. The overall score was used in the calculation of these estimates. For each squadron, the average intercorrelation among raters was computed. Two estimates of reliability were produced, first the estimated reliability for a single rater and second, the reliability for all raters. For the supervisory SARS, the average interrater

correlation across the 11 squadrons was computed to be .50. The estimated reliability for all raters was found to be .84. These data clearly demonstrate the increase in the reliability of the scores through the addition of multiple raters.

SARS Intercorrelations. The third set of analyses computed intercorrelations among the three sets of SARS scores, which are presented in Table 4. For the sake of brevity, only correlations with the overall score are presented for both the self-report and supervisory SARS. The average correlation of category scores with the overall score was computed to be .95 for the supervisory SARS and .86 for the self-report SARS, indicating a high degree of internal consistency. All correlations were statistically significant ($p<.01$) and, as seen in Table 4, the relationships among the supervisory and peer ratings were quite high. Although the correlations of the self-report SARS with the other ratings were positive, their magnitude was substantially lowered at $p < .01$.

TABLE 4. SARS INTERCORRELATIONS

	1	2	3	4
1. Supervisory SARS-Overall Score	-			
2. Peer SARS-Fighter Pilot Ability	.85	-		
3. Peer SARS- SA Ability	.87	.97	-	
4. Peer SARS-Rank Order	.87	.89	.92	-
5. Self-Report SARS-Overall Score	.50	.55	.58	.53

SARS Composite Score Development. In developing a composite SARS score, it was decided to exclude the self-report SARS for two reasons. First, the self-report SARS was significantly influenced by squadron membership. And second, only moderate correlations were found with the supervisory and peer ratings. Consequently, only the 3 peer SARS scores and the 8 supervisory SARS category scores were included in the development of a single composite score. The overall score from the supervisory SARS was excluded since, mathematically, it represented a linear combination of the category scores. A principal components analysis was performed to determine the underlying structure of these scores. The first principal component was found to account for 92.5% of the total variance of these scores, the second component 3.3%, and the remaining components less than 1% of the variance. Based upon these results, it was decided to compute composite scores based upon the first unrotated principal component due to the large amount of variance it explained. These scores were transformed to a distribution with mean of 100 and standard deviation of 20 for use as the composite SA score in subsequent analyses.

Effects of Previous Experience on Composite SARS Score. Analyses were performed to determine if the composite SA score was related to previous flight experience information. It seemed reasonable to expect that measures of experience such as flight hours, flight qualification, and combat training exercise participation should be related, to some extent, to our composite score. In fact, if such relationships were not found, it would seriously question the validity of our composite SA score. Experience factors that were analyzed included: total flight time; total flight time in the F-15; exercise participation (i.e., number attended) including Red Flag (0, 1, 2 , >3), Green Flag (0, >1) Maple Flag (0, >1) and William Tell (0, >1); air combat simulation training experience (yes/no) including the McDonnell-Douglas Advanced Air Combat Simulation (MACAIR) and the Simulator for Air-to-Air Combat (SAAC); Desert Storm experience (yes/no); and current flight qualification including whether the pilot was a Fighter Weapons School graduate (yes/no). Additionally, the effect of squadron membership was also analyzed. A one-way ANOVA was computed for each factor, except for flight time. For total flight hours and flight hours in the F-15, correlations were computed. The results are summarized in Table 5.

TABLE 5. EFFECTS OF BACKGROUND AND EXPERIENCE FACTORS ON COMPOSITE SA SCORE

	F-Ratio	*p*
Squadron	1.09	NS
Flight Qualification	143.84	<.001
Exercise Participation		
Red Flag	23.94	<.001
Green Flag	6.46	<.01
Maple Flag	4.83	<.05
William Tell	1930	<.001
Fighter Weapons Grad	47.61	<.001
Simulator Experience		
MACAIR	27.42	<.001
SAAC	5.06	<.05
Desert Storm Veteran	1.19	NS
F-15 Hours	.62*	<.001
Total Flight Hours	.40*	<.001

*Correlations

As shown in Table 5, most of the experience factors were related to the composite measure of SA. In fact, only one of the measures was not significantly related to SA--participation during Desert Storm. It should also be noted that squadron membership had no effect on the composite SA score. In all cases, the direction of the means was such that higher experience was associated with better SA scores. In fact, some of the relationships were extremely high. For example, current flight qualification accounted for 68% of the variance of the SA measure. These means are presented in Figure 1. As shown, there is a very strong relationship with flight qualification.

In the final set of analyses, a prediction equation was derived for the composite SA score using a combination of background and experience factors. A stepwise regression analysis was performed with the composite SA score as the dependent variable and those statistically significant background experience factors listed in Table 5 as the potential set of predictor variables. A "dummy variable" coding scheme was employed to enable entry of flight qualification which is a categorical variable. A four variable "best fit" prediction equation was produced with a multiple R of .85. Variables included in the equation, in their order, were flight qualification, graduation from Fighter Weapons School, participation at Green Flag and participation at Maple Flag. The overall multiple R was statistically significant ($p<.0001$) as well as the contribution of each variable within the equation ($p<.05$).

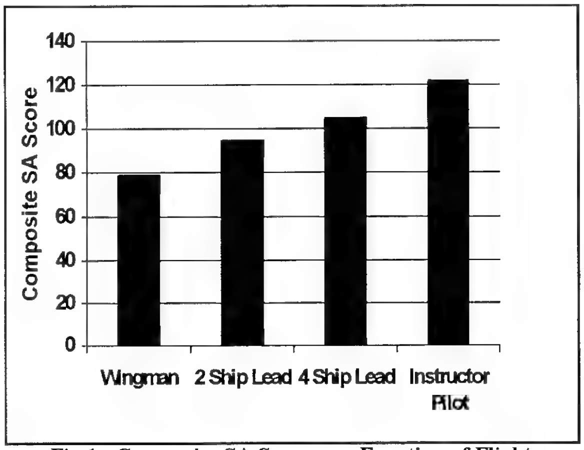


Fig 1. Composite SA Score as a Function of Flight Qualification

4 DISCUSSION

Three measurement tools were developed for assessing SA within an operational fighter environment. The primary concerns with any measurement device are its reliability, susceptibility to unwanted bias factors, and its validity.

SARS Reliability. Reliability estimates, in all cases, were quite high. Estimates of internal consistency for both the self-report and supervisory SARS exceeded .97, indicating that, whatever the 31 items might be measuring, they are indeed measuring it consistently. Of greater importance, however, are the estimates of inter-rater reliability. It was reasoned that both the reliability and validity of the criterion SA scores would be enhanced if each ratee was evaluated by multiple raters. Consequently, for the supervisory SARS, each ratee score was based on an average of from 5 to 8 raters. The results of the reliability analyses confirm the value of such an approach. The average reliability across squadrons obtained for a single rater for the supervisory SARS was marginal. However, there occurred a large increase when the average scores for all raters were used as the estimate. Although such increases in reliability from use of multiple raters seem intuitive, the performance rating literature (15) has not always produced such effects.

Interrelationships Among SARS. An analysis of interrelationships among the SARS scores produced high correlations between the supervisory and peer SARS scores, in fact, extremely high correlations. Certainly, the magnitude would not have been expected from the previous literature (15). Of greater consistency with the literature, however, were the relationships with the self-report SARS scores. Although positive correlations were obtained between the self-report SARS and the supervisory and peer SARS, their magnitudes were significantly lower. Moreover, a comparison of the overall means between the supervisory and self-report SARS revealed higher means for the self-report SARS scores, which is consistent with the previous findings of a "leniency" effect of self-ratings (15).

The high degree of consistency between the supervisory and peer SARS scores was further confirmed by the principal components analysis in which the first component accounted for over 92.5% of the total variance. The average correlation between the 8 category SARS scores and the first component score was .96. The second component accounted for an additional 3.3% of the variance and seemed to represent some unique variance associated with the peer SARS. Correlations with the component score were .34 and .33 for fighter pilot and SA ability, respectively, and .19 for the ranking. All correlations with the supervisory SARS scores were negative and most (6 of 8) were not statistically significant. Overall, these results further substantiate the high agreement between the supervisory and peer SARS score and the existence of a very large component that can account for most of the variance. Although there does appear to be a second component that is capturing some unique variance associated with the peer SARS, its size was quite small, and consequently not used as a criterion measure of SA.

Effects of Flight Experience. At the outset, we hypothesized that there would be positive relationships between flight experience and the SA criterion measure. In fact any measure that was unrelated or negatively related to flight experience would be highly suspect. The results clearly supported our hypothesis in that most of the experience data produced positive relationships with the composite SA score. The finding of positive correlations of both total flight time and time in the F-15 is consistent with the earlier literature, although the obtained correlations were higher than what has been reported in the past. The variable found most highly correlated with the criterion SA measure was current flight qualification. In fact, this variable alone accounted for nearly 68% of the criterion variance. When flight qualification was combined with other available information, a prediction equation could be developed which accounted for nearly 73% of the criterion variance, which is equivalent to a correlation of .85. These results clearly indicate that the criterion measure of SA developed for this study can be predicted reasonably well

from readily-available background and flight experience information.

Interpretation and Use. Two questions emerge from these findings. First, what is actually being measured by the SARS? And second, what are implications for use of the composite SARS score as a criterion measure? An inherent problem with most criterion measures is that they usually represent a "picture" in time (22). Within the operational fighter environment, pilots progress in a fairly "lock step" manner as they move from one flight qualification to another. Each F-15 pilot begins his career in an operational fighter squadron by completing mission qualification training. At that point he is designated a mission-ready wingman. After a certain number of hours in the jet, he becomes eligible for upgrade to 2-ship lead. If successful, he gains experience, i.e. flight hours, and eventually becomes eligible for upgrade to 4-ship lead. Within this process, a certain amount of "selection" occurs. If he is judged not to have the requisite skills for upgrade, his career as a fighter pilot will usually end and he will be reassigned. Viewed in this manner, it is not surprising that current flight qualification is highly related to our criterion measure of SA. It is clear that all raters (both supervisors and peers) were aware of each ratee's flight qualification within the squadron. Such knowledge likely provided a good frame of reference and, to some unknown extent, may have been the basis for making judgments required in the SARS. Consequently, it appears that the SARS, in large part, measures what might be termed an Air Force "management" view of fighter pilot skill, and as such, would be highly correlated with flight experience and current qualification. However, the criterion SA measure is more than "experience only" as indicated by the number of what might be termed exceptions. For example, there occurred instances in which an individual's criterion SA score was "inconsistent" with his qualification level. For example, there were IPs who received scores that were more characteristic of wingmen. And conversely, there were some wingmen and 2-ship leads that received scores much higher than their experience would suggest.

Implications for use of the composite SA score as a criterion measure are fairly straightforward. It is clear that effects due to background and flight experience must be controlled when these scores are used as criterion measures. This could be accomplished by partialling out these effects and using scores representing the residual variance as measures. Alternatively, separate analyses could be conducted for each qualification category. Regardless, the fact remains that experience accounted for a very large percentage of the variance within our criterion measure.

One other implication of these study findings should be noted. For future investigations requiring some criterion measure of performance in operational squadrons, the peer rating technique is recommended for several reasons. First, it was found to be highly correlated with the supervisory ratings. Second, and perhaps most importantly, it took considerably less time to administer. In the current investigation, the time required for subjects to complete the background questionnaire, the importance ratings, the self-report ratings, and the peer ratings was around 20 minutes. In contrast, the time required for supervisors to complete the entire package, including the detailed supervisory SARS for each pilot in the squadron, ranged from two to three hours. Given the fact that flying supervisors are usually quite busy, such a reduction in time required for data gathering could lead to considerable manpower savings and squadron acceptance.

5 REFERENCES

1. McMillan, G.R., Bushman, J. & Judge, C.L.A. (1995) Evaluating pilot situational awareness in an operational environment. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

2. Carretta, T.R. & Ree, M.J. (1995) Determinants of situational awareness in U.S. Air Force F-15 pilots. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

3. Waag, W.L., Houck, M.R., Greschke, D.A. & Raspotnik, W.B. (1995) Use of multiship simulation as a tool for measuring and training situation awareness. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

4. Carroll, L.A. Desperately seeking SA. TAC Attack (TAC SP 127-1) 1992; 32: 5-6.

5. Sarter, N.B., & Woods, D.D. Situational awareness: A critical but ill-defined phenomenon. Int. J. Aviat. Psychol. 1991; 1:45-57.

6. Endsley, M.R. Predictive utility of an objective measure of situational awareness. Santa Monica Ca: Human Factors Society, Proceedings of 34th Annual Meeting, 1990: 41-45.

7. Fracker, M.L. Measures of situation awareness: Review and future directions. Wright-Patterson AFB OH: Armstrong Laboratory, 1991; AL-TR-1991-0128.

8. Youngling, E.W., Levine, S.H., Mocharnuk, J.B., & Weston, L.M. Feasibility study to predict combat effectiveness for selected military roles: Fighter pilot effectiveness. St Louis MO: McDonnell Douglas Astronautics Co, 1977; MDC E1634.

9. Strawbridge, D., & Kahn, N. Fighter pilot performance in Korea. Chicago, IL: Institute for Air Weapons Research, 1955; IAWR 55-10.

10. Torrance, P.E., Rush, C.H., Kohn, H.B., & Doughty, J.M. Factors in fighter-interceptor pilot combat effectiveness. Lackland AFB TX: Air Force Personnel and Training Research Center, 1957; AFPTRC-TN-57-23.

11. DeLeon, P. The peacetime evaluation of the pilot skill factor in air-to-air combat. Santa Monica, CA: The Rand Corporation, 1977; R-2070-PR.

12. Lepley, W.M. (Ed.) Psychological research in the theaters of war. Washington D.C.: US Government Printing Office, 1947; AAF Psychol. Prog. Res. Report No. 17.

13. Shannon, R.H., & Waag, W.L. Toward the development of a criterion for fleet effectiveness in the F-4 fighter community. Pensacola, FL: Naval Aeromedical Research Laboratory, 1972; NAMRL-1173.

14. Shannon, R.H., & Waag, W.L. The prediction of pilot performance in the F-4 aircraft. Pensacola, FL: Naval Aeromedical Research Laboratory, 1973; NAMRL-1186.

15. Landy, F.J., & Farr, J.L. Performance rating. Pschol. Bull. 1980; 87: 72-107.

16. Jenkins, J.G., Ewart, E.S., & Carroll, J.B. The combat criterion in naval aviation. Washington D.C.: National Research Council on Aviation Psychology, 1950.

17 . Houck, M.R., Whittaker, L.A., & Kendall, R.R. An information processing classification of beyond-visual-range intercepts. Williams AFB, AZ: Armstrong Laboratory, 1993; AL/HR-TR-1993-0061.

18. Houck, M.R., Thomas, G.S., & Bell, H.H. Training potential of multiplayer air combat simulation. Santa Monica CA: Human Factors Society, Proceedings of 33rd Annual Meeting, 1989: 1300-1304.

19. Houck, M.R., Thomas, G.S., & Bell, H.H. Training evaluation of the F-15 advanced air combat simulation. Williams AFB, AZ: Armstrong Laboratory, 1991; AL-TP-1991-0047.

20. Thomas, G.S., Houck, M.R., & Bell, H.H. Training evaluation of air combat simulation, Williams AFB, AZ: Air Force Human Resources Laboratory, 1990; AFHRL-TR-90-30.

21. James, J.M., Sanderson, P.M., & Seidler, K.S. (1990). SHAPA version 2.0: Instruction manual and reference (EPRL 90-16/M). Urbana, IL: University of Illinois.

22. Austin, J.T., & Villanova, P. The criterion problem: 1917-1992. J. Appl. Psychol. 1992; 77:836-874.

SITUATION AWARENESS: IMPACT OF AUTOMATION AND DISPLAY TECHNOLOGY

Christopher D. Wickens
University of Illinois at Urbana-Champaign
Institute of Aviation
Aviation Research Laboratory
Savoy, IL 61874 USA

1 SUMMARY

In this paper we first offer two compatible definitions of situation awareness, present an information processing model of how it is maintained and lost, and discuss different techniques by which it is measured. Causes for the loss of situation awareness related to low visibility, dense airspace, free flight, and automation are briefly described, and then problems related to automation-induced situation awareness loss with the flight management system are discussed in detail. These problems are related to a poor mental model, high system complexity, removal of the pilot from the control loop, and inadequate displays. The next section of the paper focuses on display technology that has been used to support situation awareness of aviation hazards (traffic, weather, terrain). We discuss research issues related to electronic map scale, rotation, and dimensionality as these influence flight path guidance and hazard awareness. It is concluded that the more egocentric levels of these dimensions that do the best job of supporting flight guidance do not necessarily support situation awareness. The final section briefly discusses the less well researched area of task awareness.

2 INTRODUCTION

In 1992, a Thai Airlines jet, after executing a missed approach to the Kathmandu Airport, apparently became disoriented with regard to direction, and flew into a 6000 meter mountainside. Also in 1992, the pilots of an Airbus A320 Jet reached an excessive rate of decent on their approach to Strassbourg, eventually crashing the aircraft into the ground at a vertical decent rate of 3000 feet/min. Indications are that they may have confused two autopilot mode settings on the vertical decent mode, setting 3000 feet/min, rather than the (apparently) intended 3.0 degree flight path angle. In the case of each of these tragic accidents, it is apparent that the pilots did not fully understand what was "going on," either in the aircraft systems themselves (the Strassbourg incident), or in the world beyond the aircraft (the Kathmandu crash). That is, they were **unaware of the situation**. Furthermore, in at least one of these cases (the Strassbourg crash), it appears that the pilots were unaware that they were unaware.

These two cases, and hundreds of other accidents and incidents that have been reported, illustrate instances of a breakdown of **situation awareness (SA)**, the focus of this symposium, I will start by providing some definitions and conceptual foundations for the concept of situation awareness. I will briefly describe some of the psychological components underlying the maintenance, loss and recovery of situation awareness, and I will then address issues of its measurement. I will then focus on two current critical issues; how developments in automation may contribute to the loss of situation awareness, and how advances in display technology may contribute to its restoration.

3 DEFINING SITUATION AWARENESS

The definition of situation awareness can be approached from two useful directions. One is via the formal definition of its components. The second is through the "consensus" definition that has emerged from the community of researchers and pilots who have been most concerned with the concept.

The American Heritage Dictionary defines a **situation** as "the **place** at which something is located" and "the combination of circumstances at a given **moment**." Hence, these definitions emphasize both place and time. The more psychological construct of **awareness**, is defined by Yates [1] as that which is "voluntarily reportable through language, pressing a button, sorting and classifying and so forth." Hence, we see a definition that incorporates voluntarily reportable information bearing on the certain aspects of the intersection of space and time.

This formal definition is relatively consistent with the more specific consensus definition that emerges from the "user community." Although the number of separate definitions offered by workers and practitioners within their community is large, I will resist the temptation to offer still another definition, and instead rely upon the work of Dominguez [2], who has carefully reviewed the many studies in this area, and the many definitions offered, to derive a "consensus" definition. In the following I highlight in italics, a few words of my own, which I have added to the Dominguez [2] definition:

"Situation awareness is the continuous extraction of environmental information *about a system or environment*, the integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception, anticipating *and responding to* future events."

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

This definition thus highlights the three processing components of situation awareness identified by Endsley [3]: the perception of information, the comprehension of that information, and the prediction of its future implications. I have added the explicit distinction between system and environmental awareness, because this distinction is so critical to aviation; what is happening within the aircraft versus what is happening concerning the aircraft's relation to surrounding hazards. I have also added the characteristic of **responding,** because good situation awareness will not just allow the pilot to anticipate an event, but it will help provide the pilot with the capacity to **respond appropriately** to events that may or may not be anticipated; thus the pilot with good situation awareness may not anticipate the failure of the automated system (presumably a VERY low frequency and therefore unexpected event); but this pilot will at least have the capacity to respond appropriately if the system does fail. Thus, we may think of good situation awareness as supporting the **potential** to perform effectively in unexpected circumstances. In this regard, there is a close analogy between situation awareness and workload [4,5,6]. Workload is defined as inversely related to a "reserve capacity," and, like situation awareness, it also describes the **potential** of the operator, in this case, to handle unexpected increases in task demands.

Conceptually, we may represent our concept of situation awareness as in Figure 1, highlighting within the cylinder, those aspects of the current and future situation that are part of the mental picture; that is, those aspects which can be readily brought to mind when needed. The "breadth" of the cylinder represents, conceptually, the domain of space (here shown as 2D rather than 3D) about which information is known or is predictable. This breadth diminishes in the past because of breakdowns in memory. It diminishes in the future because of limits in our predictive abilities and the uncertain behavior of the world.

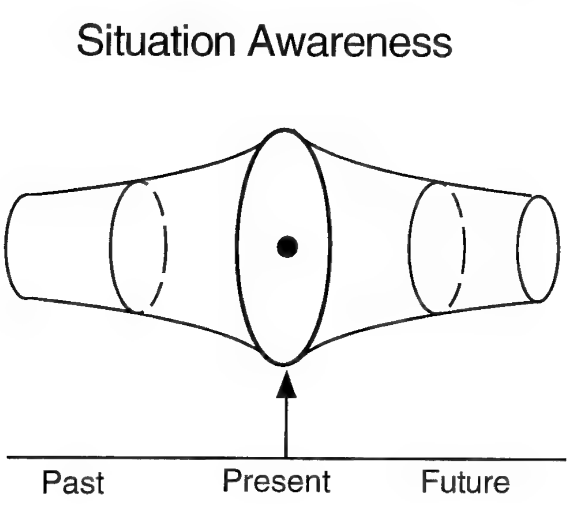


Figure 1: Conceptual representation of situation awareness in space and time.

In formatting our definition of situation awareness, it is also important to specify what our definition is NOT, in order to provide some restrictions and specificity. First, situation awareness is NOT long term memory, knowledge, or skill, primarily because this knowledge does not typically evolve or change with the frequency of situational changes. As we see below, long term knowledge may be necessary to support good situation awareness; but highly skilled pilots, possessing extensive knowledge may be very poor at updating their mental picture, if the displays are poor, if they are looking in the wrong place, or if they are distracted.

Second, situation awareness is NOT the same as performance (although performance measures may sometimes be necessary to measure situation awareness). Thus, for example, a pilot flying with a flight director can generate very accurate performance in adhering to a trajectory; yet may not have the slightest idea of where his aircraft is, with respect to the ground, other aircraft, or hazardous weather; thus, his situation awareness is low. As we will see later, we have found that the displays that support the best **performance** in terms of flight path guidance, are actually poorest at supporting awareness of the location of surrounding terrain features [7]. Conversely, the passenger in an aircraft jumpseat who is physically disabled, may have extremely good situation awareness, by attending carefully to all channels of information; yet be totally unable to "perform" the flying task (i.e., control the aircraft).

Third, I emphasize that situation awareness is not an "all or none" thing, such that you have it or you don't. Like all psychological constructs, situation awareness is a matter of degree. There may be different ranges of space over which it is maintained, and different degrees of precision or accuracy with which it is maintained at a given location of space. There may be different ranges into the future with which it is maintained, and people may have varying degrees of confidence in their own knowledge. All of these "varying" features contribute to a view that situation awareness is not a simple state.

4 A PROCESSING MODEL OF SITUATION AWARENESS

Figure 2 presents a model of what we propose to be the perceptual and cognitive processes involved in the maintenance of situation awareness, and the influences on those processes. Situation awareness itself may be seen to lie within the shaded region. At the "top" of the model are represented attention and perception, accessing information which is typically derived from displays, but also from observations and voice communications [8]. Perceptual information is also closely related to information maintained in working memory, just below. Working memory is conscious, shortlived, of limited capacity, and requires active

rehearsal (demanding limited resources) for its maintenance. Clearly working memory contents represent a component of situation awareness. However, as we have noted, the notion of "availability" also allows information that is not in the momentary consciousness of working memory, to be considered part of situation awareness. Here then we include the **retrievable** portions of long term memory (pertaining to the evolving situation). If a pilot has just "set" an autopilot to a particular mode, or an air traffic controller has just positioned an aircraft at a fix, these operator's do not need to continuously rehearse these facts (at the expense of other operations), in order for them to be considered as part of situation awareness. These items of information can, and presumably will be retrieved when needed, and the failure to retrieve them is considered a clear breakdown in situation awareness, as a recent airplane crash at the Los Angeles Airport demonstrated when a local controller forgot that an aircraft had been positioned on an active runway.

been forgotten. The figure also shows, to the right, as if on a time axis, a component designated **prediction,** hence incorporating Endsley's [3] critical third level of situation awareness; the ability to use present information (typically perceived from displays), in conjunction with a mental model of how the world typically works (in long term memory), in order to "compute" an expectation of future state. Such computation is typically resource intensive, and will break down if other tasks compete. Hence, we may describe prediction as heavily dependent upon the resources of working memory [9].

As shown around the margins of the figure, situation awareness is influenced by several factors that will be discussed in more detail below. **Displays,** if effectively constructed, can support situation awareness by providing information in an intuitive, simple, and comprehensive format and directing attention to the appropriate place at the appropriate time. **Perception**

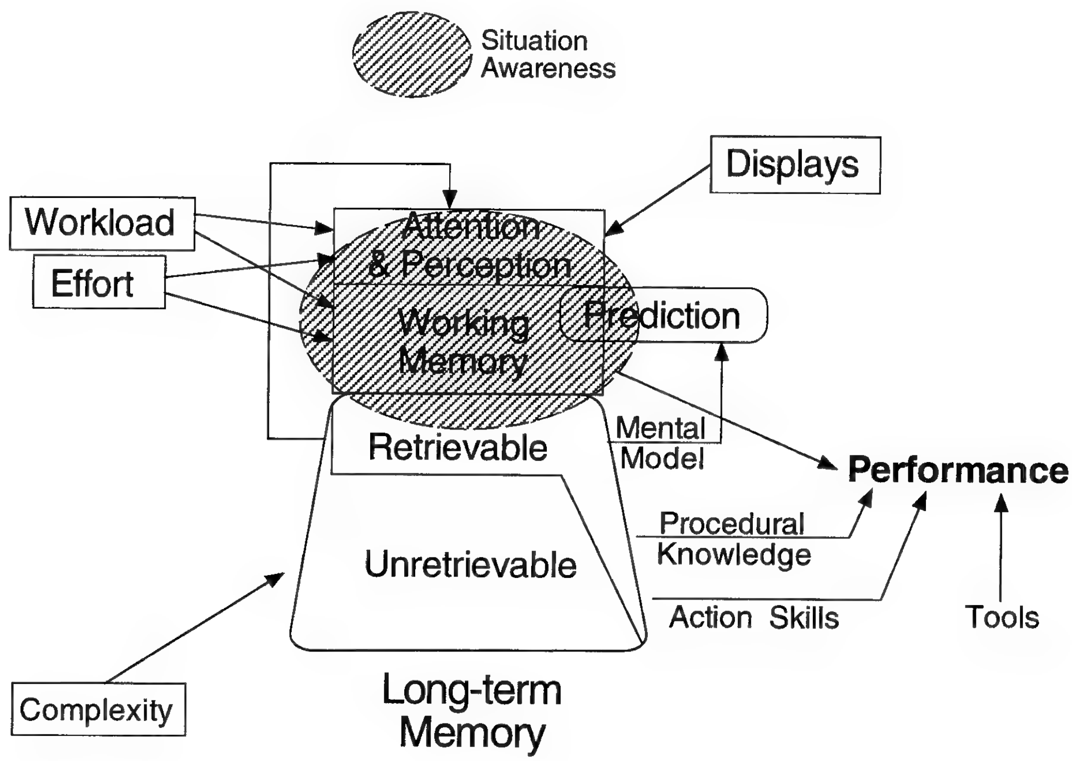


Figure 2: An information processing model of situation awareness, its relation to performance, and the external influences on its maintenance.

As shown in the figure, long term memory contains a large amount of information that cannot be easily retrieved. Some of this information may pertain to aspects of the dynamic situation that have been forgotten, whereas much else may consist of facts and pieces of semantic or structural information that has

of events to gain situation awareness is also supported by long term memory, since the skilled user of any system has an effective "schema" with which to more easily absorb and interpret new information. Extensive **workload** can degrade situation awareness, either by competing with working memory resources, or by competing with the perceptual resources necessary to

aggregate new information. Conversely, **effort**, not diverted to other tasks but allocated to the situation, may serve to enhance situation awareness. As we see below, one source of this effort is the requirement for the operator to be an active participant in making decisions about the system; decisions that lead to the changes in system state.

In addition to supporting perception, long term memory also may support the understanding of system state (i.e., a well-developed **mental model**), which is important for prediction in establishing situation awareness. The mental model is, of course developed through training. However, it may also be degraded by high levels of system complexity (or, alternatively, be harder to overcome by training because of these high levels). This we will see is a potential negative feature of the high levels of complexity of automated flight management systems [10,11].

Finally, the figure indicates that situation awareness supports, but is not identical to performance, since the latter also depends upon procedural knowledge, action skills and the physical tools to carry out the task. Hence, a key facet of situation awareness is how to measure its contributions to performance, uncontaminated by the contributions of these latter "action" factors. We turn now to this issue of measurement.

5 MEASUREMENT OF SITUATION AWARENESS

Situation awareness can be measured subjectively or objectively, and the latter by explicit, implicit, current or retrospective techniques. Each of these have their costs and benefits. **Subjective measures** entail asking either the operator, or those peers and supervisors of the operator, if the operator has situation awareness. In some circumstances investigators have found that peer and supervisor ratings of a pilot's situation awareness, can provide an adequate prediction of subsequent performance [12].

Yet as has often been noted, particularly in the area of workload assessment, subjective measures have their limitations. With regard to situation awareness, for example, the problem with asking the operator is that people don't always "know what they don't know" [10], and are at times inherently overconfident about their knowledge state, feeling that they know more than they do [9]. People sometimes use their long term knowledge to construct and report a plausible inference of what is going on, rather than base this report on actual evolving perceptual information.

Because of these concerns about voluntary, subjective (usually verbal) reports, **performance measures** must also be considered as indices. An **implicit** situation awareness measure considers the question: if the operator did NOT have situation awareness, could we discriminate his performance on a natural part of the task, from the operator who did have situation awareness? As we have noted, many aspects of routine flying performance cannot make this discrimination. For example, the pilot using the flight director can fly the plane just as well with as without situation awareness. However, the pilot who must suddenly decide if a controller's clearance to a new flight path is safe, will respond very differently with, than without situation awareness [13]. Hence, the latter form of measurement (the decision to reject an unsafe clearance), provides an effective implicit measure.

A limitation of implicit measures of situation awareness is that, in general, they must be fairly narrow in scope. For example, assessing the response to a single in-flight unexpected event (which if dealt with effectively indicates high situation awareness), will yield assessment of only a small component of the total situation confronting the pilot just prior to the event.

A somewhat wider scope of situation awareness can be assessed by **explicit measures** which interrupt the natural flow of the task, to "ask" the operator about the situation; "where are other aircraft in the sky?," "who is friend and who is foe?," "which way is north?," "where is the closest mountain?," etc. [14,7]. These techniques have the advantage of assessing directly the information that is needed and can be employed to assess a fairly broad range of issues. But they have the possible disadvantage that they disrupt the flow of the "natural task" (although this disruption may not be too much of a problem [3,15]. More seriously, if the same explicit measures are assessed repeatedly, they could lead the operator to perform the task in a manner that is quite different from the way it might naturally be performed. For example, if one is repeatedly asked about the location of terrain features, one might spend an inordinate amount of time trying to memorize those locations. Thus, there is a danger that explicit measurements may change the measurement environment to one that is less generalizable to the real world conditions of flight. To prevent this from occurring then, it is important that the pilot be aware of the wide scope of possible questions that **could** be asked, so as to make it unfeasible to prepare for any particular query type.

Finally, explicit measures may be **retrospective**. After a mission has been completed, operators can be asked to describe their flight path, or reconstruct the position of hazards or sequence of events within the environment. This approach, of course, has the advantage of being completely non-obtrusive of the ongoing task, since that task will have been completed prior to the measurement. The technique is limited, however, because a failure of long term memory recall (which would be used to infer poor situation awareness), does not necessarily indicate that the information was not in awareness at some earlier time. The technique is also limited because sometimes past experience leads people to reconstruct plausible sequences of events, on the basis of what usually happens (a mental script) rather than what actually did happen.

There are, of course, other measures that can correlate with situation awareness because they may be indices of the kinds of things that enhance or degrade it. For example, workload measures may reveal excessive workload, that could be expected to degrade situation awareness [14,4], or measures of visual scanning could be used to infer that information sampling was restricted, in a way that can degrade situation awareness. While such measures are important, and may sometimes be necessary, one must be aware that they are not sufficient, and should never be confused with measures of situation awareness, itself.

Finally, because of the nature of situation awareness as a mental construct, it is my firm belief that no single measure can be fully adequate in all circumstances. Converging evidence from a variety of implicit and explicit measure, buttressed by other measures, must be used to infer the loss of situation awareness, and the effectiveness of different techniques for its restoration, the two areas that we shall now address.

6 LOSS OF SITUATION AWARENESS

Situation awareness only becomes an issue in aircraft safety when it is lost. Such a loss may result from systematic transient or chronic deficiencies on the part of the pilot (momentary inattention, loss of vigilance, fatigue or poor training), or because certain flight conditions inhibit its maintenance. We focus here on the latter categories of which we can identify four of particular relevance:

1. Low visibility, in instrument meteorological conditions (IMC) degrades, and sometimes eliminates important information (e.g., accurate view of the terrain below, or precise visualization of target aspects in the forward view).

2. A denser airspace, the result of increasing air traffic demands, will make the maintenance of situation awareness more difficult to achieve whether in IMC or VMC, simply because of the shear number of elements that must be monitored. This problem will be equally shared by the pilot and the air traffic controller.

3. The potential of the airspace to shift to a "free flight" regime, in which increasing responsibility for maintaining safe separation is shifted from the ground (in ATC) to the air (by the pilot) will, if it occurs, impose heavy demands on situation awareness. In parallel with this shift will evolve more complex "envelopes" of safe separation around each aircraft, that are defined in terms of the minimum time to contact, rather than fixed metric separations (i.e., the "hockey puck"). While minimum time envelopes are more flexible and optimal for defining the most efficient use of limited airspace at a given safety level, they are also considerably more complex to visualize, and hence, will challenge both the pilot's and the controller's ability to maintain awareness.

4. As we discuss in detail below, automation will, and has already decreased the operators situation awareness in many circumstances [3,11]. We address this last issue focusing in particular on the case study of the modern flight management system, or FMS [10], although the issues may characterize any of a number of other aviation automation systems, such as those envisioned for automated scheduling, conflict detection and resolution in air traffic control [16].

7 FLIGHT MANAGEMENT SYSTEM AND THE LOSS OF SITUATION AWARENESS

The flight management system or FMS consists of a sophisticated collection of sensors and controls that will allow the pilot to program in advance various higher levels of goals for the aircraft to attain (e.g., intersect a particular approach path, change flight levels, airspeeds and headings contingent upon other conditions), and then allow the automation to carry out the necessary tasks

7.1 The Problem

While the FMS usually carries out its task silently, correctly and efficiently, there are nevertheless a non-trivial number of exceptions. In fact, a frequently quoted paraphrase of pilot's responses to many advanced automation systems is: "what did it do?, why did it do it?, and what will it do next?" [17,18,19]. These words are verbalizations of "automation induced surprises," reflecting a lack of situation awareness which has been documented systematically by a series of experimental investigations carried out by Sarter and Woods [20,10; see also 21], and supported by aircraft incident analyses [17,18], as well as reconstruction of several recent accidents [19].

In the experimental studies, carried out in air transport simulators with line pilots fully qualified to fly automated aircraft, the pilot participants answered questions regarding how they thought the FMS worked, and also had to respond to unexpected and unusual configurations of the aircraft as it flew under FMS control (i.e., implicit measures of situation awareness). While most pilots were effective in setting up and using the FMS for normal operations, a substantial number revealed inadequate situation awareness under conditions when the system would be unexpectedly configured in an unusual, but not impossible, state. These configurations might result from an erroneous pilot input, from the need to respond to unexpected external events (e.g., a missed approach), or from a possible failure of some aspect of the automation. Under these circumstance, a substantial number of pilots simply failed to understand what the FMS was doing and why; they were surprised by its behavior in a way that would make questionable their ability to respond appropriately. Four sources of such surprises can be represented in the context of Figure 3.

7.1.1 Poor mental model. Pilots may be well trained in how to set up the FMS to accomplish particular goals, but they may have little training on the details of **how** the FMS actually implements the controls to carry out those goals, nor are pilots often allowed to exercise

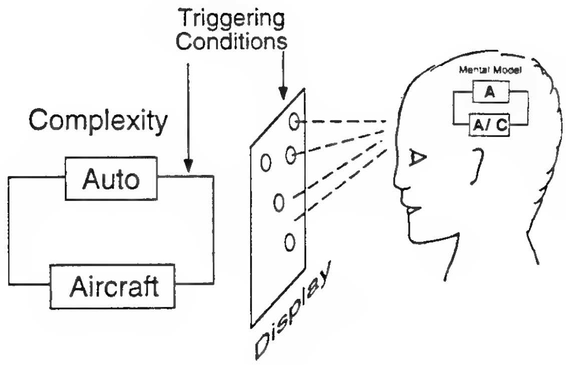


Figure 3: The influences on situation awareness of automated aircraft control.

and refine those mental models by observing the consequences in unusual configurations. In the absence of good knowledge of how a system works, it is hard to maintain a good mental picture of how it is configured at any given point in time, and to use anticipation as a guide for where to attend, as the events unfold over time.

7.1.2 System complexity. In the efforts to serve the pilot with maximum flexibility, designers have created a large number of different options; there are, for example, at least five different modes whereby altitude can be changed. Yet the more complex the system, the more difficult it is to keep track of its changing states [23], particularly if the changes are not well-supported in training and by displays (see 7.1.4).

7.1.3 Participatory mode. Automation of control systems (like an aircraft) places the human operator in a participatory mode that is outside of the control loop, passively monitoring its inputs and outputs. This participatory mode has two consequences to human performance [3,11]. The first of these is the now well-documented phenomenon of **complacency** [24], sometimes referred to as overtrust [11,25]. If the automated system works well, the operator may trust its performance so well that the operator fails to monitor its inputs and outputs closely, and hence, will fail to **detect** (or detect slowly) circumstances in which the automation or the system itself fails [26], particularly if this failure is subtle and "graceful."

The second consequence, related to but distinct from the first, may be described in terms of the **memory benefits** of **active choice**. There is by now a well-documented literature regarding the better memory for changes in state when the human operator has been an active agent in bringing about those changes, rather than a passive participant witnessing another agent making the same changes. That is, the act of **doing** (choosing) facilitates the memory of the **done** (chosen) [27,28,29,16].

Both of these factors together can serve to reduce the pilot's situation awareness of the setting of automated modes, the status of inner loop flight variables, and the status of key "triggering conditions" or parameters that will cause other components of the automated system to initiate an action (e.g., reaching a particular altitude). Collectively, they produce what we refer to as a state of "out-of-the-loop unfamiliarity" or OOTLUF. Furthermore, both factors (but complacency in particular) will inhibit the pilot's development of a mental model, in a way to contribute still further to the loss of situation awareness.

7.1.4 Poorly integrated displays. The information by which the state of the aircraft's automation systems is presented to the pilot in relatively scattered, unintegrated form, distributed across a **mode control panel** just below the windscreen, the **primary flight display**, and the **control display unit** of the flight management computer, to the pilot's side [10]. Such information is represented in various forms, from alphanumeric codes of what modes are in effect, to lights that may or may not be illuminated. Missing from the picture is an integrated **spatial representation** of what the plane is doing and will be doing in the airspace, as time proceeds, along with salient perceptual signals that will help the pilot anticipate and attend to the location and identity of significant automation-induced events. That is, a display to support **automated situation awareness.**

7.2 The Solutions

Our discussion of the problems in the previous section makes fairly implicit many of the solutions that can be implemented to address an automation induced situation awareness loss [30]. First, more active "exploratory" training, for example, of the whole range of FMS activities in unusual as well as predictable circumstances would be of great value [22]. Second, it seems reasonable that some efforts to simplify the complexity of the algorithms would be useful, perhaps by reducing the number of options, even if this does reduce the flexibility and power of the FMS.

Third, it seems clear that efforts to develop integrated temporal/spatial displays can be of great value [31], displays that can provide explanation of what is happening and salient guidance for when events will be expected. This issue we treat in depth in the following section. Fourth, and perhaps most challenging are the solutions to the OOTLUF problem. How does one keep the pilot sufficiently "in the loop" so that awareness of system changes is maintained via active choice, without defeating entirely the purposes of automation, which are to remove the pilot sufficiently FROM the loop that the workload required to achieve accurate flight is not excessive. The solution would seemingly be to seek the appropriate **level** of required pilot intervention to either make, or at least approve

automation recommendations of different trajectory choices [27].

8 DISPLAYS TO SUPPORT HAZARD SITUATION AWARENESS

Unfortunately, very little research has been carried out on designing display formats to support automation relevant situation awareness, as discussed in the previous section. However, a more positive picture can be painted regarding research and development of electronic displays for airborne **hazard** situation awareness, in which the location of hazards including weather, terrain or other aircraft is portrayed. Such displays **must** by definition be electronic, because the evolving aspect of **situations** requires dynamic updating capabilities. Early developments in these areas include the horizontal situation display (electronic map) in many commercial and military aircraft, and the Cockpit Display of Traffic Information (CDTI) developed by NASA [32,33], which became a precursor to the present TCAS alert system [34]. Both of these display prototypes, and others we discuss below have in common their dynamic characteristics, and the fact that they present a wider geographical range of information than that which is minimally required to fly the anticipated flight path, existing in a spatial "cone" forward of the aircraft (Figure 4, Panel A).

While there have been many efforts to develop different advanced display concepts for global situation awareness, there have been far fewer that have systematically set out to graphically **evaluate** such concepts with pilot performance in controlled studies comparing the advanced concepts against their more traditional counterparts. Such evaluations are necessary in order that designers can ascertain whether the display offers an improvement, and can determine what may be the psychological mechanism that is responsible for the improvement (or possible cost) to pilot performance and situation awareness. We review below the conclusions of studies that have exercised such control with regard to three critical features of aeronautical hazard display design: the "breadth of coverage" or **scale** of the display, the **frame of reference** of the display, and the **dimensionality** of the display. As shown in Table 1, each of these dimensions may be characterized by an endpoint that is either **egocentric**, (characteristic of the view a pilot actually has of the real world in VRF flight) or **exocentric** (more characteristic of a stabilized "God's eye" view of the world). The distinction between ego and exocentric displays is important, because of emerging evidence that the more egocentric displays, focusing on a 3D view looking directly forward from the cockpit, tend to support better flight path control [35,7]. Yet the question we must ask at this symposium is whether such a gain in guidance performance must

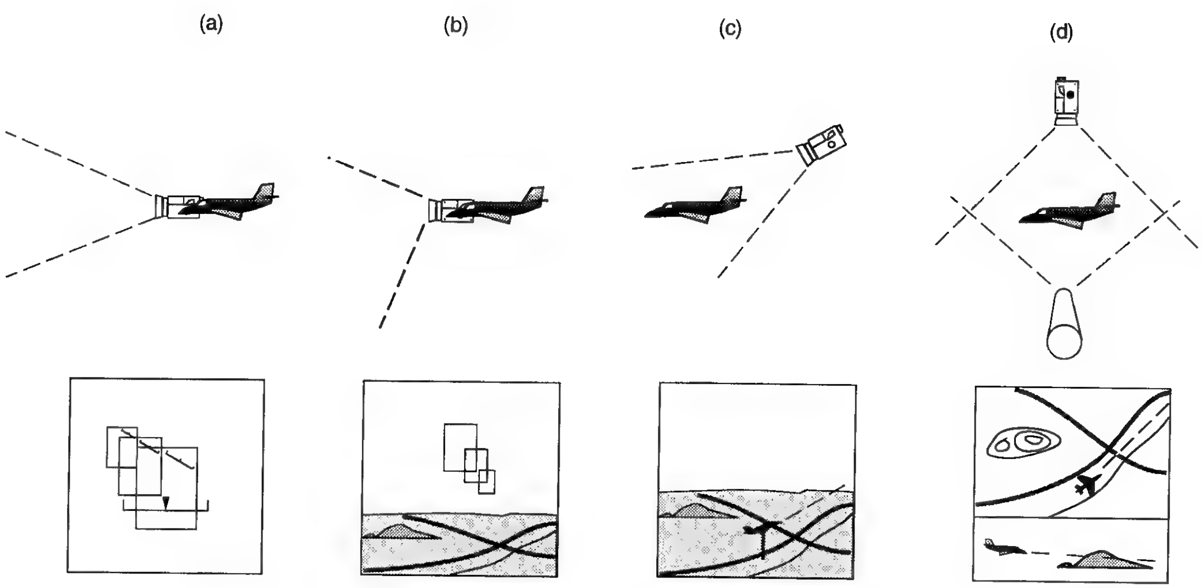


Figure 4: Different frames of reference for the depiction of aircraft hazard information. In the top row, the "camera" depicts the view point of the display that would be situated in the cockpit. The schematic scene that might be viewed by the pilots is depicted in the bottom view.

be purchased at the inevitable price of a loss in situation awareness. In this final section we try to answer this question by examining each of these three dimensions in turn.

Table 1

	Egocentric		Exocentric
Scale of Display	"Zoom In" (Telephoto)	-	"Zoom Out" (Wide Angle)
Frame of Reference	Rotating (Track Up)	-	Fixed (North Up)
Dimensionality	3D-Perspective	-	2D-Planar (Top & Profile View)

8.1 Display Scale

The issue of display scale is the simplest to address, and will be considered first. Quite simply, the optimal scale on a display, whether 2D or 3D, fixed or rotating, depends upon the distance away from ownship for which situation awareness information is needed. The greater this distance, the greater is the region that should be portrayed. The cost of "wide angle" large scale displays portraying a great distance is also evident. These will produce a loss of spatial precision in judging where things are.

8.2 Map Rotation

This design dimension characterizes whether electronic maps (i.e., displays to characterize the location of things in geographical space) should rotate in a track up fashion, according to the momentary heading of the aircraft, or should remain fixed in what is generally a north up orientation. Aside from the technological factors (e.g., more dynamic imagery updating is required with the rotating maps), the issue has important pilot performance implications. By now a fairly extensive series of studies has indicated that guidance or flight path control is better when the pilot is supported by rotating maps. This advantage results primarily because fixed (north up) maps force pilots to engage in time and effort consuming **mental rotation** when they are flying on southerly legs, in order to translate between the axis of display and the axis of control [36,37; Figure 5). That is, under these conditions, what is right on the map is leftward in terms of its control implications.

It appears, however, that the issue of fixed versus rotating maps is far less well-resolved with regard to situation awareness than it is with guidance. This is, in part, because the knowledge required of situation awareness may itself be either expressed in ego-referenced terms ("an aircraft is at the 4:00 position") or world-referenced terms ("a severe thunderstorm is 270 degrees (due west) from your current location").

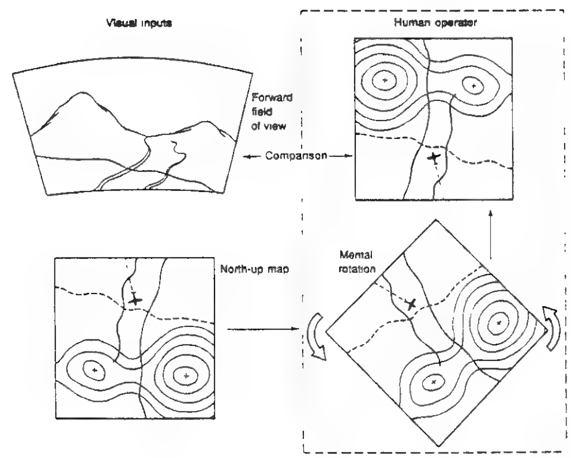


Figure 5: Representation of the cognitive process of mental rotation, when flying south with a north-oriented map.

Thus, for situation awareness assessment tasks that depend more on ego-referenced knowledge, rotating map displays have generally proven better [13,37,38,39]. However, for those situation awareness tasks requiring access to knowledge in a world-referenced frame, like indicating absolute bearings [39,40] or finding locations of things on a map [41], fixed maps appear to be superior. We have also noted however, that in tasks in which active navigation (i.e., "flying") is extensively involved, the workload costs of mental rotation imposed by fixed map displays, may "spill over" and consume resources otherwise available for other aspects of performance, like learning about the environment [42] or controlling altitude [39,43].

While the above results speak directly to concurrent measures of situation awareness, whether implicit [37,13] or explicit [39,43], there is some evidence that retrospective measures of long term memory for geographical features reveal a slight, but not consistent advantage for the fixed map format. This is presumably because the consistent manner in which the geographical information is presented allows a more stable mental representation of the world to be formed [37,44,39,43].

In sum, it our data would suggest that, if a single display were to be available to serve all tasks (i.e., both guidance and situation awareness), with greatest benefits and fewest costs, then map rotation should be the display of choice, although we would strongly advocate the availability of fixed map options.

8.3 Dimensionality

Three-dimensional displays have often been proposed as advantageous for maintaining situation awareness [45], and in military tactical combat situations, the added "intuitive" and pictorial realism offered by such displays may indeed provide such support. However, 3D displays for aviation are encumbered by at least two kinds of deficiencies that must be carefully considered in terms of their potential negative impact on situation awareness, an impact that may neutralize or even reverse their advantages.

First, the costs of typical "highway in the sky" displays, such as that shown in Figure 4a and Figure 6 is simply that it presents a fairly narrow field of view of the forward cone. One solution to this deficiency is to change the design parameters to present a wider field of view (Figure 4b). However, this option either consumes greater amounts of display real estate, or forces a distortion (compression) of real space that can severely disrupt situation awareness (i.e., knowing where things are; [46]). A second solution to the first problem is to pull the viewpoint of the display away

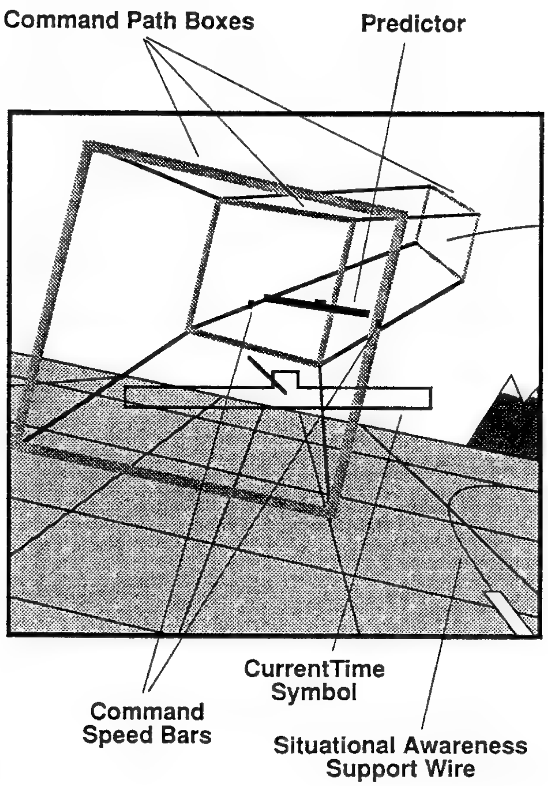


Figure 6: Example of a fully ego-referenced display shown in Figure 4(a), developed by Haskell and Wickens (1993).

from the location of own-ship, as shown in Figure 4 Panel C, so that a much wider region of space can be perceived in a less distorted fashion. However, this feature produces another source of perceptual distortion related to perceptual "compression" that is involved whenever egocentric judgments are made from an exocentric viewpoint [46,47,48,49].

A third solution is to replace or augment flat panel displays positioned in the forward field-of-view with **omnidirectional** displays. These may be either helmet-mounted displays or auditory localization displays [50]. While both technologies have limitations, both can present accurate and intuitive information from an egocentric reference frame regarding the bearing of hazards in the full 360° sphere of space around the pilot.

The second deficiency of 3D displays, characteristic of either ego- or exocentric displays involves the ambiguity of position estimate along the line of sight or viewing vector of the display. Without including some sort of augmented perceptual cues, there are an infinite number of points in 3D space that can be collapsed onto any 2D projection surface. Hence, any 3D display will be vulnerable to this loss of precision, as long as it represents objects at different distances from the observer (or distances that are unknown to the observer). And since much of situation awareness requires knowing the position of objects in space, the potential cost to 3D displays imposed by this factor is evident.

Against these costs may be arrayed the full benefits of a 2D planar display **suite** that presents at least a top down and a profile view of the airspace surrounding ownship (Figure 4d). An unlimited range of space on all sides of the aircraft can be presented, and precise and unambiguous information regarding distances and angles separating the aircraft from other hazards can be viewed. The primary cost of such a suite is the evident **visual scanning**, that may be needed to combine the two views, and the **cognitive effort** required to integrate the two frames into a single integrated representation of the aircraft and hazards in 3D space. However, this effort may not be extensive if the situation awareness task is not one that requires integration between vertical and lateral planes [51], but can deal with information in each plane sequentially.

Research that has examined 2D versus 3D displays to support situation awareness has, indeed revealed this tradeoff of influences, suggesting that there is no consistent benefit of one format over the other. Three conclusions that do emerge from this research appear to be as follows:

1. 3D tunnel in the sky displays, in which the viewpoint of the display corresponds to the position of the pilot (i.e., fully egocentric; Figure 4a and Figure 6), are superior for flight path control [7,35]. Their integrative capabilities support the pilot's need to integrate present and future lateral and vertical deviations of the aircraft, an integration which is

required given the cross coupling of the axes in aircraft flight dynamics. Furthermore, the correspondence of the display viewpoint with the axis of control and rotation is a natural and intuitive form of display-control compatibility. However, as noted above, these displays do little for situation awareness because of their narrow field of view.

2. The distortion, caused by increasing the field of view (minifying) of 3D ego-referenced displays (Figure 4b), an increase which is necessary to portray a wide region of space, appears to be sufficiently harmful so as to seriously disrupt situation awareness of where things are [7].

3. 3D exocentric displays (Figure 4c) are neither consistently better nor worse than their 2D counterparts (Figure 4d), in supporting situation awareness, reflecting the tradeoff of factors discussed above. Thus, we have found that, relative to their 2D counterparts, such 3D displays inhibit the judgment of where things are on both the vertical axis [39,43,52], and the lateral axis [53]. However, this accuracy loss is not always found [7] and, because of their integrated characteristics which reduces time-consuming visual scanning, 3D displays may sometimes allow judgments to be made more rapidly [39,7].

As a consequence of these findings, a full evaluation of whether or not 3D displays will help situation awareness must be based upon a careful evaluation of the costs of lost precision with 3D viewing (some "holistic" tasks, like understanding the general shape of the data base may not require much precision), relative to the need for integration between vertical and lateral axes, better supported by 3D displays; and these factors in turn must be weighed by the range of space around and away from the aircraft that must be presented, and the costs of visual scanning between separate panels in a 2D suite. Because these factors trade off in complex ways, the choice between the three formats shown in Figure 4b,c,d cannot easily be ascertained in advance.

9 TASK SITUATION AWARENESS

Accidents have sometimes resulted from the pilot's failure to perform critical tasks; not necessarily because the pilots have been so overloaded that they tried and could not complete the tasks, but rather because the need to perform those tasks somehow dropped from the pilot's awareness. Examples of such tasks may include checking altitude [54] lowering a landing gear, or setting up an automated device. These breakdowns in **cockpit task management** [55,56] may be described in terms of situation awareness, if one envisions a "task space" defined by axes of priority and time, and populated by tasks that are either queued to be performed, or have just been performed, as time moves by. Far less is known about how situation awareness of this task space is maintained and lost, and about appropriate displays to support it (other than checklists; [57]. We can anticipate, however, that such a task space will become more complex in multioperator environments, in which there is shared responsibility for certain tasks. Such environments include also those characterizing flexible or adaptive automation [58]. Indeed the issue of task management is an important component of the emerging concept of **crew resource management** in the multioperator cockpit [59], and may be anticipated to play a very critical role as issues of task responsibility may shift between ground and air through the possible introduction of "free flight" in the future air space.

Ironically the checklist, itself, designed as a form of task situation awareness support, may evolve to an automated level that can sacrifice the very task awareness it was designed to maintain [57]. This is because of the possible implementation of varying levels of automation in checklist maintenance; from automation "reminders" to pilots that certain checklist tasks have not been completed (preserving high task awareness), to full automation assumption of responsibility for carrying out those tasks (danger of OOTLUF). Here again, design attention must be (and is being) given to consider the optimal levels of automation that will reduce human error, but still retain operator task awareness [57].

10 CONCLUSIONS

We have seen, regrettably, that automation and display configurations that are best suited for some aspects of routine performance may not necessarily be those that best support situation awareness. With regard to automation, high levels of autopilot control can minimize flight path deviations and maximize energy management and fuel economy, but will degrade the pilot's knowledge of aircraft state. With regard to displays, those that offer high resolution guidance information regarding the forward cone of vision (or prescribe appropriate commands) will do so at the expense of broader awareness of hazards in the full 360° sphere of airspace around and away from the pilot. Even in the case of the yet little investigated area of task awareness, it is likely that effective displays of tasks to be performed and their priority will consume valuable real estate that might be served by other displays.

However, this state of affairs need not be treated too pessimistically. It is not the case that the designer **must** choose one configuration or the other (guidance support or situation awareness support). Rather, the challenge to the researcher is to seek compromise levels of automation [3,27] or display configurations [7], that can provide an adequate level of support for both kinds of tasks. This search, coupled with careful consideration by the system analyst of the relative importance of the two tasks for safety and mission success, should lead to configurations that may indeed be able to create "the best of both worlds" or at least, nearly the best.

11 ACKNOWLEDGMENTS

This research was sponsored by a grant NASA NAG 2-308 from NASA Ames Research Center. Vernol

Battiste and Sandra Hart have served as the technical monitors on this grant. The author would like to acknowledge the helpful comments of Mica Endsley, Nadine Sarter, and Michael Vidulich on an earlier draft of this paper, and the contributions of Chia-Chin Liang to the review of the literature.

13 REFERENCES

1. Yates, J., "The Content of Awareness is a Model of the World," Psychological Review, 92, 1985, pp. 249-284.

2. Dominguez, C., "Can SA be defined?", *in* Vidulich, M., Dominguez, C., Vogel, E., and McMillan, G., "Situation Awareness: Papers and Annotated Bibliography (U)," Interim Report AL/CF-TR-1994-0085, Armstrong Laboratory, Air Force Materiel Command, Wright-Patterson Air Force Base, OH, 1994.

3. Endsley, M.R. "Automation and Situation Awareness," *in* M. Mouloua and R. Parasuraman (Eds.), "Automation and Human Performance," Hillsdale, NJ, Lawrence Erlbaum Associates, 1994.

4. Wickens, C.D., "Workload and Situation Awareness: An Analogy of History and Implications," Insight, 14, 4, 1992b, pp. 1-3.

5. Endsley, M.R., "Situation Awareness and Workload: Flipsides of the Same Coin," *in* "Proceedings of the 7th International Symposium on Aviation Psychology," Dept. of Aviation, Ohio State University, Columbus, OH, pp. 1993, 906-911.

6. Roscoe, A.H. (Ed.), "The Practical Assessment of Pilot Workload," AGARDograph No. 282 (AGARD-AG-282), NATO AGARD, Neuilly Sur Seine, France, 1987.

7. Wickens, C.D. and Prevett, T., "Exploring the Dimensions of Egocentricity in Aircraft Navigation Displays: Influences on Local Guidance and Global Situation Awareness," Journal of Experimental Psychology: Applied, 1, 1995, 110-135.

8. Midkiff, A. and Hansman, R.J., "Analysis of the Importance of Party Line Information in ATC Operations," MIT Aeronautical Systems Laboratory Report ASL-92-2, Massachusetts Institute of Technology, 1992.

9. Wickens, C.D., "Engineering Psychology and Human Performance, 2nd ed.," New York, Harper Collins, 1992a.

10. Sarter, N.B. and Woods, D.D., "Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System," International Journal of Aviation Psychology, 4, 1, 1994, pp. 1-28.

11. Wickens, C.D., "Designing for Situation Awareness and Trust in Automation," *in* G. Johannsen (Ed.), "Proceedings of the International Federation of Automatic Control (IFAC) Conference on Integrated Systems Engineering," Oxford, UK, Pergamon Press, 1994, pp. 77-82.

12. McMillan, G.R., "Report of the Armstrong Laboratory Situation Awareness Integration (SAINT) Team (Briefing Transcript)," Interim Technical Report AL/CF-TR-1994-0085, "Situation Awareness: Papers and Annotated Bibliography (U)," Armstrong Laboratory, Air Force Materiel Command, Wright-Patterson AFB, OH, 1994, pp. 37-47.

13. Mykityshyn, M.G., Kuchar, J.K., and Hansman, R.J., "Experimental Study of Electronically Based Instrument Approach Plates," International Journal of Aviation Psychology, 4, 2, 1994, pp. 141-161.

14. Fracker, M.L. and Davis, S.A., "Explicit, Implicit, and Subjective Rating Measures of Situation Awareness in a Monitoring Task," Final Report AL-TR-1991-0091, Armstrong Research Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, OH, 1991.

15. Vidulich, M., Dominguez, C., Vogel, E., and McMillan, G., "Situation Awareness: Papers and Annotated Bibliography (U)," Interim Report AL/CF-TR-1994-0085, Armstrong Laboratory, Air Force Materiel Command, Wright-Patterson Air Force Base, OH, 1994.

16. Hopkin, D.V., "Human Performance Implications of Air Traffic Control Automation," *in* M. Mouloua and R. Parasuraman (Eds.), "Human Performance in Automated Systems: Current Research and Trends," Hillsdale, NJ, Lawrence Erlbaum Associates, 1994, pp. 314-319.

17. Wiener, E.L., "Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft," NASA Contractor Report No. 177528, NASA-Ames Res. Ctr., Moffett Field, CA, 1989.

18. Rudisill, M., "Flight Crew Experience with Automation Technologies on Commercial Transport Flight Decks," *in* M. Mouloua and R. Parasuraman (Eds.), "Human Performance in Automated Systems: Current Research and Trends," Hillsdale, NJ, Lawrence Erlbaum Associates, 1994, pp. 203-211.

19. Dornheim, M.A., "Dramatic Incidents Highlight Mode Problems in Cockpits," Aviation Week & Space Technology, Jan. 30, 1995, pp. 57-59.

20. Billings, C.E., "Toward a Human-Centered Aircraft Automation Philosophy," The International Journal of Aviation Psychology, 1, 4, 1991, pp. 261-270.

21. Sarter, N.B. and Woods, D.D., "Pilot Interaction with Cockpit Automation: Operational Experiences with the Flight Management System," International

Journal of Aviation Psychology, 2, 4, 1992, pp. 303-321.

22. Irving, S., Polson, P., and Irving, J.E., "A GOMS Analysis of the Advanced Automated Cockpit," *in* "CHI-94 Conference Proceedings," New York, Association of Computing Machinery, 1994.

23. Wohl, J.G., "Connectivity as a Measure of Problem Complexity in Failure Diagnosis," *in* "Proceedings of the 27th Annual Meeting of the Human Factors Society, Santa Monica, CA, Human Factors Society, 1983, pp. 681-683.

24. Parasuraman, R., Molloy, R., and Singh, I.L., "Performance Consequences of Automation-Induced Complacency," International Journal of Aviation Psychology, 3, 2, 1993, pp. 1-23.

25. Lee, J. and Moray, N., "Trust, Control Strategies, and Allocation of Function in Human-Machine Systems," Ergonomics, 35, 1992, pp. 1243-1270.

26. Wickens, C.D. and Kessel, C., "Failure Detection in Dynamic Systems," *in* J. Rasmussen and W. Rouse (Eds.), "Human Detection and Diagnosis of System Failures," New York, Plenum Press, 1981.

27. Endsley, M.R. and Kiris, E.O., "The Out-of-the Loop Performance Problem and Level of Control in Automation," Human Factors, in press.

28. Williams, H.P. and Wickens, C.D., "Electronic Map Interpretation in a Dual-Task Context," ARL-91-2/NASA-91-2, Savoy, IL, University of Illinois Institute of Aviation Technical Report, Aviation Res. Lab., 1991.

29. Cowan, N., "Evolving Conceptions of Memory Storage, Electronic Storage, Selective Attention, and Their Mutual Constraints Within the Human Information Processing System," Psychological Bulletin, 104, 2, 1988, 163-191.

30. International Federation of Automatic Control Panel on Automation in Commercial Aviation I, Baden-Baden, Germany, IFAC, Sept. 1994, pp. 65-82.

31. Hahn, E.C. and Hansman, R.J., "Experimental Studies on the Effect of Automation on Pilot Situational Awareness in the Datalink ATC Environment," SAE AEROTECH Conference and Exposition, Warrendale, PA, Society of Automotive Engineers, Inc., 1992.

32. Abbott, T.S., Moen, G.C., Person, L.H., Jr., Keyser, G.L., Jr., Yenni, K.R., and Garren, J.F., Jr., "Flight Investigation of Cockpit-Displayed Traffic Information Utilizing Coded Symbology in an Advanced Operational Environment," NASA TP 1684, Hampton, VA, NASA Langley Res. Ctr., 1980.

33. Ellis, S.R., McGreevy, M.W., and Hitchcock, R.J., "Influence of a Perspective Cockpit Traffic Display Format on Pilot Avoidance Maneuvers," AGARD Conference Proceedings No. 371: Human Factors Considerations in High Performance Aircraft (AGARD-CP-371), Neuilly Sur Seine, France, 1984, pp. 16-1/16-9.

34. Chappell, S.L., "Human Factors Research for TCAS," *in* "Proceedings of the Royal Aeronautical Society Conference on the Traffic Alert and Collision Avoidance System (TCAS) - Its Development, Testing, Installation and Operational Use," 1990, pp. 6.1-6.19.

35. Haskell, I.D. and Wickens, C.D., "Two- and Three-Dimensional Displays for Aviation: A Theoretical and Empirical Comparison," International Journal of Aviation Psychology, 3, 2, 1993, pp. 87-109.

36. Aretz, A.J. and Wickens, C.D., "The Mental Rotation of Map Displays," Human Performance, 5, 1992, pp. 303-328.

37. Aretz, A.J., "The Design of Electronic Map Displays," Human Factors, 33, 1, 1991, pp. 85-101.

38. Marshak, W.P., Kuperman, G., Ramsey, E.G., and Wilson, D., "Situational Awareness in Map Displays," *in* "Proceedings of the 31st Annual Meeting of the Human Factor Society, Human Factors Society, Santa Monica, CA, 1987, pp. 533-535.

39. Wickens, C.D., Liang, C.-C., Prevett, T., and Olmos, O., "Egocentric and Exocentric Displays for Terminal Area Navigation," *in* "Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society Annual Meeting," Human Factors and Ergonomics Society, Santa Monica, CA, 1994.

40. Hofer, E., "Evaluation of Rotation of Electronic Approach Plates in a Full Mission Simulator," Ph.D. Dissertation in progress, 1995.

41. Harwood, K. and Wickens, C.D., "Frames of Reference for Helicopter Electronic Maps: The Relevance of Spatial Cognition and Componential Analysis," International Journal of Aviation Psychology, 1, 1991, pp. 5-23.

42. Williams, H., Wickens, C.D., and Hutchinson, S., "Realism and Interactivity in Navigational Training: A Comparison of Three Methods," *in* "Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society Annual Meeting," Human Factors and Ergonomics Society," Santa Monica, CA, 1994.

43. Liang, C.C., Olmos, O., and Wickens, C.D., "Perspective Electronic Map Evaluation in Visual Flight. *in* R. Jensen (Ed.), "Proceedings of the 8th International Symposium on Aviation Psychology," Columbus, Ohio, Dept. of Aviation, Ohio State University, 1995.

44. Faye, E. and Wickens, C.D., "Strategies for Display Integration in Navigational Guidance and Situation Awareness," University of Illinois Institute of Aviation

Technical Report, ARL-95-4/NASA-95-1, Aviation Research Laboratory, Savoy, IL, 1995.

45. Reising, J.M. and Mazur, K.M., "3-D Displays for Cockpits: Where They Payoff," *in* "Proceedings of the SPIE/SPSE Symposium on Electronic Imaging Science and Technology," Santa Clara, CA, Feb., 1990.

46. McGreevy, M.W. and Ellis, S.R., "The Effect of Perspective Geometry on Judged Direction in Spatial Information Instruments," Human Factors, 28, 1986, pp. 439-456.

47. Ellis, S.R., McGreevy, M.W., and Hitchock, R.J., "Perspective Traffic Display Format and Airline Pilot Traffic Avoidance," Human Factors, 29, 1987, pp. 371-382.

48. Wickens, C.D., Todd, S., and Seidler, K.S., "Three-Dimensional Displays: Perception, Implementation, and Applications," State-of-the-Art Report, CSERIAC SOAR 89-001, Armstrong Research Laboratory, Wright-Patterson AFB, OH, 1989.

49. Perrone, J.A., "Visual Slant Underestimation: A General Model," Perception, 11, 1982, pp. 619-625.

50. Begault, D.R. and Wenzel, E.M., "Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces," The International Journal of Aviation Psychology, 2, 1, 1992, pp. 1-22.

51. Wickens, C.D. and Carswell, C.M., "The Proximity Compatibility Principle: Its Psychological Foundation and Its Relevance to Display Design," Human Factors, in press.

52. Endsley, M.R., "Measurement of Situation Awareness in Dynamic Systems," Human Factors, 1995.

53. Boyer, B.S., Campbell, M., May, P., Merwin, D.H., and Wickens, C.D., "Three-Dimensional Displays for Terrain and Weather Awareness in the National Airspace System," *in* "Proceedings of the 1995 Annual Meeting of the Human Factors and Ergonomics Society," Santa Monica, CA, Human Factors and Ergonomics Society, 1995.

54. Wiener, E.L, "Controlled Flight into Terrain Accidents: System-Induced Error," Human Factors, 19, 1977, pp. 171-181.

55. Funk, K. "Cockpit Task Management: Preliminary Definitions, Normative Theory, Error Taxonomy, and Design Recommendations," The International Journal of Aviation Psychology, 1, 4, 1991, pp. 271-286.

56. Raby, M. and Wickens, C.D., "Strategic Workload Management and Decision Biases in Aviation," International Journal of Aviation Psychology, 4, 3, 1994, pp. 211-240.

57. Palmer, E. and Degani, A., "Electronic Checklists: Evaluation of Two Levels of Automation," *in* "Proceedings of the Sixth Symposium on Aviation Psychology," Columbus, OH, The Ohio State University, Department of Aviation, 1991.

58. Mouloua, M. and Parasuraman, R. (Eds.), "Human Performance in Automated Systems: Current Research and Trends," Hillsdale, NJ, Lawrence Erlbaum Associates, 1994.

59. Wiener, E.L., Kanki, B.G., and Helmreich, R.L. (Eds.), "Cockpit Resource Management," San Diego, Academic Press, 1993.

TACTICAL COCKPITS - FLAT PANEL IMPERITIVES

Eugene C. Adam
MDC Distinguished Fellow
Advanced Crew Systems
McDonnell Douglas Aerospace
PO Box 516
St. Louis, MO 63166, USA
(314) 234-1060

ABSTRACT

A cockpit revolution is in the making. Many of the much ballyhooed, much promised, but little delivered technologies of the 70's and 80's will finally come of age in the 90's just in time to complement the data explosion coming from sensor and processing advances. Technologies such as helmet systems, large flat panel displays, speech recognition, color graphics, decision aiding and stereopsis, are simultaneously reaching technology maturities that promise big payoffs for the third generation cockpit and beyond.

The first generation cockpit used round dials to help the pilot keep the airplane flying right side up. The second generation cockpits used Multifunction Displays and the HUD to interface the pilot with sensors and weapons. What might the third generation cockpit look like? How might it integrate many of these technologies to simplify the pilots life and most of all: what is the payoff? This paper will examine tactical cockpit problems, the technologies needed to solve them and recommend three generations of solutions.

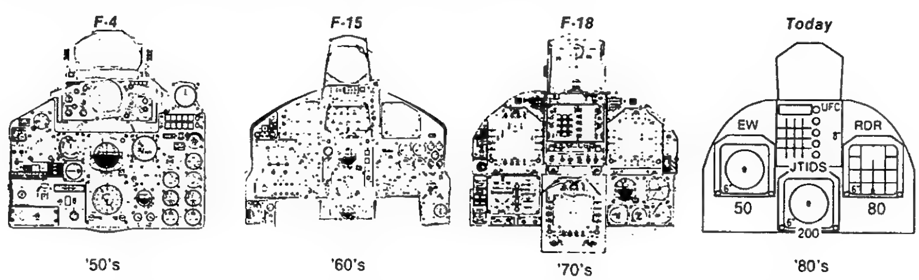
INTRODUCTION

Never has the cockpit designer had such a rich selection of emerging technologies from which to choose. But in these austere times, this treasure trove of technologies is under severe pressure to pay its way in combat kills, safety, or survivability. Therefore, each technology needs to be evaluated on the basis of which problem it solves and the cost effectiveness of the solution.

Before examining these new technologies it might be useful to first examine today's cockpits to see where we stand.

As shown in Figure 1, the analog cockpit of the two-place F-4 Phantom was followed by the HUD/CRT/Analog cockpit of the one-place F-15 Eagle which gave way to the HUD/multifunction display (glass) cockpit of the dual mission, one-place F/A-18 Hornet. Most of the western fighters built since that time use similar cockpit schemes: 1) a Head-Up Display, 2) Some Multi-Function Displays, 3) An Up-Front Control and 4) Hands on Throttle and Stick (HOTAS).

Cockpits have progressed from "steam gauges" to multipurpose displays.



However: The greatest challenge facing today's cockpit designer is to provide the pilot with the necessary Situation Awareness (SA) to be effective 'n combat. Todays cockpits have difficulty providing that SA because:

- Over 70% of Panel is inflexible
- Only 10 - 20% of Panel provides combat information
- Displays are too small to overlay Radar/NAV/EW/JTIDS on a map
- Display technology is stagnant because of low funding
- Pilot has no Head-Out Information except in the area of the HUD

GP03-0573-2-R

Figure 1. From Round Dials to Multifunction Displays. Where Do We Go From Here?

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

TWO PROBLEMS

There are two major problems with applying today's cockpit to tomorrow's sensor/mission environment: 1) today's pilot spends more time managing and integrating sensors than executing Tactics and, 2) Useful combat information is available on only 20% - 30% of the instrument panel.

Fiddling and Flying - The first problem requires the pilot to fiddle around with a host of sensors and try to mentally integrate the data from the three primary ones while flying. Radar, EW, and data link are presently displayed on three different displays, on three different range scales with two or three different "ownship" locations. In the past, this has not been an overriding problem because radar search volumes were small and they generally tracked only a few targets, EW systems were inaccurate and full of false alarms and thus largely ignored, and JTIDS/Data Links were aboard very few aircraft. This will however not be the case in the 21st Century. Sensor search volumes will increase at least one order-of-magnitude, EW accuracies will improve and data links will be common. These factors will greatly impact the pilot's ability to remain the "sensor manager/integrator" and have time left over for tactics execution.

Unproductive Space - The second problem, that of inefficient use of the instrument panel space is a straight geometry equation. The average instrument panel is roughly 18" high by 24" wide or about 400 square inches. Using (3) 5" or 6" CRT's yield a total display area of 75 to 108 square inches. Therefore, on average, 70 to 80% of the instrument panel is inflexible, devoid of combat data and unable to contribute to the fight, or bombs on target.

Since hostile contact generally averages only 30 seconds to 2 minutes the pilot has to cope with unfused data on small displays on only a fraction of the instrument panel in a time-critical, high-stress, high-g environment. Not a good formula for making "everybody an Ace".

In combat, the pilot is in the aircraft to make good tactical decisions and execute them. Everything else is secondary. However, the correctness of tactical decision-making is directly proportional to the Situation Awareness (SA) of the pilot.

SITUATION AWARENESS (SA)

So, what is SA, what is it all about? It's simply KNOWING WHAT'S GOING ON SO YOU CAN FIGURE OUT WHAT TO DO! Where are the friendlies, bogies, SAM's and unknowns with respect to my flight? What are their intentions, my intentions and my options? It's obvious that present cockpits, by separating primary sensor data, on different range scales with different "ownship" positions do not give the pilot the SA required to achieve the exchange ratios necessary to win against superior numbers of equivalent quality targets.

The Big Picture - As shown in Figure 2, SA is a two-fold problem: Global and Tactical. Global SA (the Big Picture) generally covers the non-visual spherical world at ranges from 0 to 200 miles. Most often a plan view SA is best, with your ownship position decentered because of higher interest and lethality in the forward hemisphere. However, even in a low-intensity conflict, the 100 mile range display could contain hundreds of graphic elements such as unfriendly surface and airborne threats, friendly surface and airborne elements, unknowns, navigation paths, map and symbolic data. Separate, small displays are no match for this complexity.

The Little Picture - Tactical SA covers close-in visual air-to-air and air-to-surface combat and visual navigation. M on N combat is the arena where man and machine are taxed to their limits. For equivalent machines, the SA acted upon by the eye, brain, hands and feet is the primary determinant of "who shoots" and "who chutes".

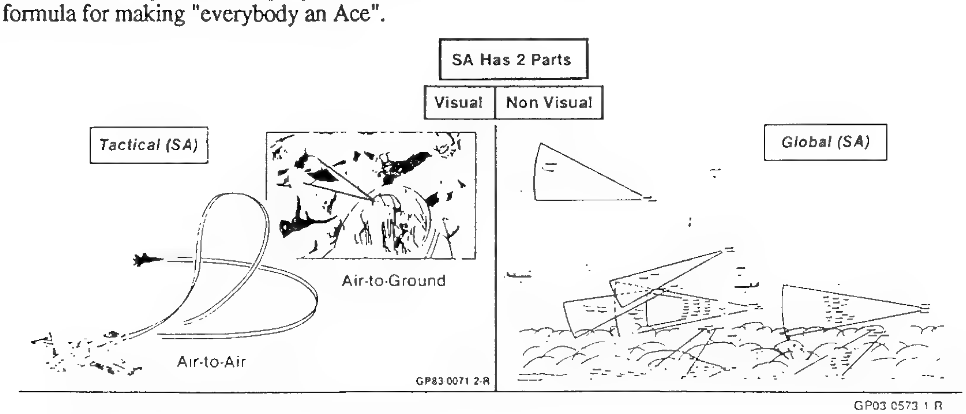


Figure 2. Situational Awareness – "Knowing What's Going on So You Can Figure Out What to Do"

TACTICAL SA SOLUTION

The tactical SA problem is best solved by a helmet system that: 1) TRACKS the pilot's head position and slaves sensors and weapons to the helmet line-of-sight, 2) DISPLAYS combat and flight information on the helmet visor.

Integrated Helmet System - MCAIR and Kaiser Electronics IRADs have designed, built, simulator tested and flown an Integrated Helmet Mounted Display and Sight (HMDS) System called "Agile Eye" (TM) which can increase visual exchange ratios by a factor of 2:1 over a Head-Up Display. The "Agile Eye" is a totally integrated helmet sight and display that has the following features:

a) A HUD type display on the visor,
b) Lighter than present helmets,
c) Improved CG,
d) Improved crash protection,
e) No visual obstructions,
f) Less aero lift during ejection,
g) Improved sound reproduction/ and attenuation.

The "Agile Eye" Helmet uses readily available off-shelf technology cleverly integrated into a pilot centered design that improves every physical and performance characteristic of today's flight helmet. It offers fields-of-view and stroke/raster capabilities that match present day HUD's but with the advantages of off-axis weapon use, three quarters the system cost, two times the reliability and the added safety of attitude and other flight data available at all times, and at all sight angles. All of these features are packaged in a low-bulk, handsome design as pictured in Figure 3.

"Agile Eye" Payoff - In A/A: Faster visual lock-ons, simultaneous AIM-7 and AIM-9 launches, target handoffs to wingman, better attitude awareness at all times. In A/G: off-boresight target designations, offset NAV waypoint updates, target handoffs to wingman. As shown in Figure 4, MCAIR F-15 simulator evaluations using TAC pilots/aggressors/scenarios showed a 2:1 exchange ratio improvement with the "Agile Eye" HMDS over the HUD.

Helmet Systems - The Linchpin - We are convinced that helmet systems are the key to future cockpit improvements; they increase a pilot's performance and free-up panel space.

WHAT IS THE PILOT'S PROBLEM?

1) Next generation sensors, such as RADAR, EW and JTIDS will provide 100's of pieces of information 2) current display technology limits CRT sizes to 5, 6 or 7 inches square, 3) small displays require separation of primary sensors such as RADAR, EW, JTIDS and NAV - leaving the pilot to mentally integrate and fuse this data during the stress of combat.

Picture This! - Three different sensors on three separate displays on three different range scales with "ownship" in three different locations; a formula for confusion. Larger displays solve that problem by fusing all sensor data to a common range scale and coordinate system and overlaying it on a map.

What is the Hardware Problem? - CRT's using a scanning beam naturally grow dimmer as they are made

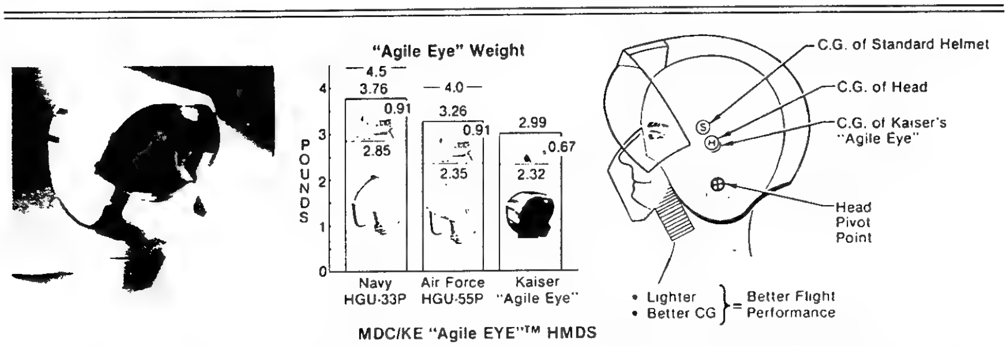


Figure 3. "Agile Eye" a HUD-on-the-Head Without Penalties to the Pilot

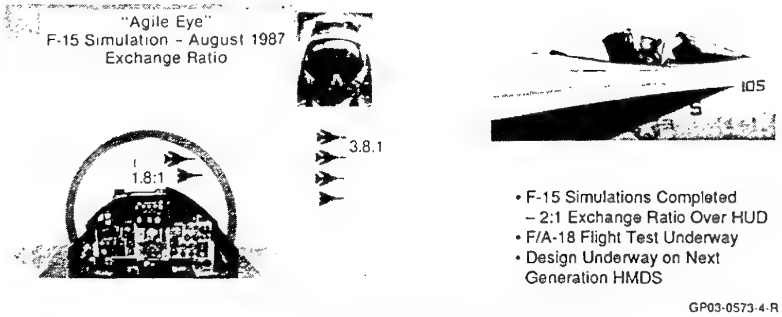


Figure 4. "Agile Eye" Doubles Exchange Ratios

larger which is unacceptable in a high ambient cockpit. Flat panel displays using matrix addressed pixels do not have this problem, but the technology and infrastructure need R&D funds before they can seriously challenge the CRT.

GLOBAL SA SOLUTION

The beyond-visual-range Situational Awareness solution requires the "fusion" of RADAR, EW, JTIDS navigation and map on a large display. This would allow the pilot to look at a single source to "get the Big Picture".

As shown in Figure 5, display size growth has not kept pace with computer and sensor technology because of the lack of serious research and development on CRT alternatives. A two-step solution offers the most cost and schedule effectiveness. In the near term, we must first develop larger, new technology displays on which to display the situation to the pilot. We must then reconfigure the HUD to provide the room to mount this display in the cockpit. In the far term we must develop new, flat-panel matrix technologies that provide display surfaces of 10 to 15 times what is available using today's CRT technology.

COCKPIT 2000: A NEAR TERM SOLUTION

Helmet systems such as "Agile Eye" are essentially a HUD-on-the-head which allows us to reduce the physical size of the aircraft HUD sufficiently to provide room for a 10" x 10" Global Situation Display. This display is a compromise between being large enough to fuse RADAR, EW and JTIDS on a single touch sensitive surface, but yet small enough to leave room for adjacent 5" or 6" auxiliary displays.

As shown in Figure 6, Cockpit 2000 has about 2X the display area of current fighters and differs from today's cockpit in two important aspects: 1) A helmet sight and display provides all normal HUD functions on the helmet visor with the added benefit of off-axis target designation, 2) The 10" x 10" Global Situation Display is larger and more productive than any three, small multifunction displays.

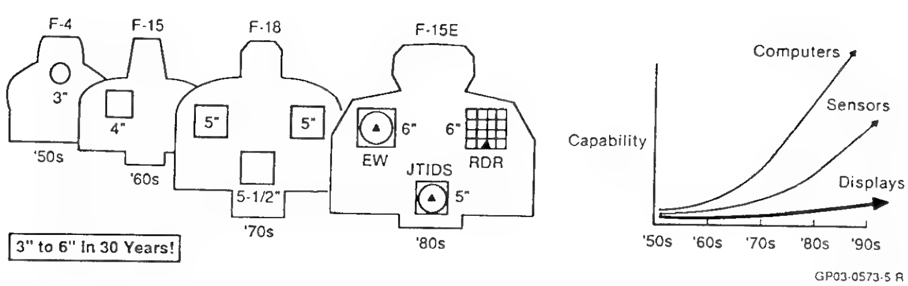


Figure 5. Present Evolution of Displays Not Keeping Up With Computers and Sensors

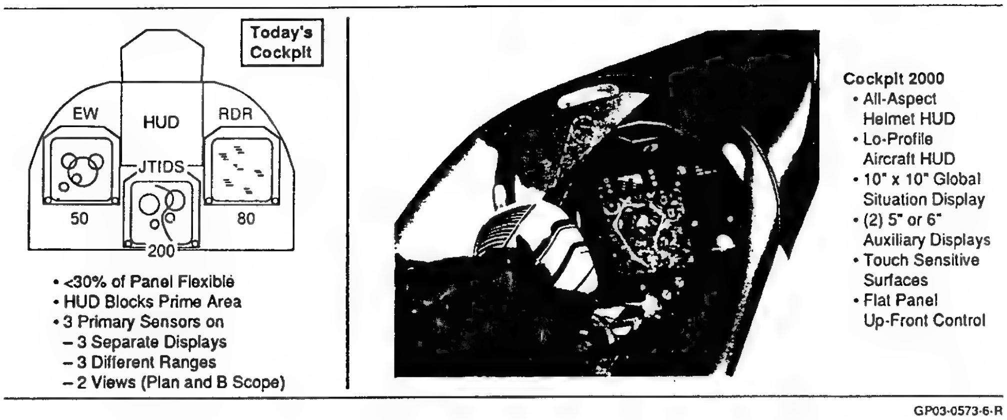


Figure 6. Cockpit 2000 Solves the Two Most Pressing Cockpit Problems, Tactical and Global SA

BIG PICTURE: A LONGER TERM SOLUTION

An increase in display technology R&D will eventually provide flat, matrix display panels with large surface areas, high brightness, high resolution and long life. As depicted in Figure 7, these large displays will provide 10 times the display area of today's CRTs allowing plan and perspective views, split screen, and movable inserts. A Helmet Sight and Display, voice command and touch sensitive surface will provide pilot interface with the weapon system. In short, the Big Picture provides the pilot with full control over the configuration and content of almost 400 square inches of display surface to match the mission-moment-of-interest whether it be air-to-air, air-to-surface, Navigation, TF/TA, or System Status. Manned Simulations have shown a 100% increase in the situational awareness of pilots using the Big Picture over those using a conventional 2 or 3 small MFD (CRT) cockpit.

Display Technology - The CRT has reigned supreme as the display device of choice for almost 100 years, with continuous evolutionary brightness, resolution, reliability and color improvements over that time. In fact, the huge CRT infrastructure and its good performance has stifled any real competitive technology investments until recently.

There are three large markets for a CRT replacement: 1) HDTV promises displays sizes of 2-5 times present CRT devices with the desire to "hang it on the wall" like a picture. 2) Portable PC's up through work stations desire high-resolution, full color, small bulk and for portable applications, low-power consumption. 3) Military and Aerospace all share a similar problem; too much data on too small a CRT surface. Larger displays are required to solve this problem but the bright sunlight conditions in aircraft must also be met which essentially dooms the CRT.

All three of these applications, and their commercial profit potential are giving a massive push to flat panel technologies. The next three years will see a R&D investment in flat panels of at least three times the total CRT alternative investments for the last 30 years. Unfortunately, the U.S. investment is roughly 5% of the worldwide investment, hence our commercial possibilities are few and our defense needs may well be supplied by offshore manufacturing facilities.

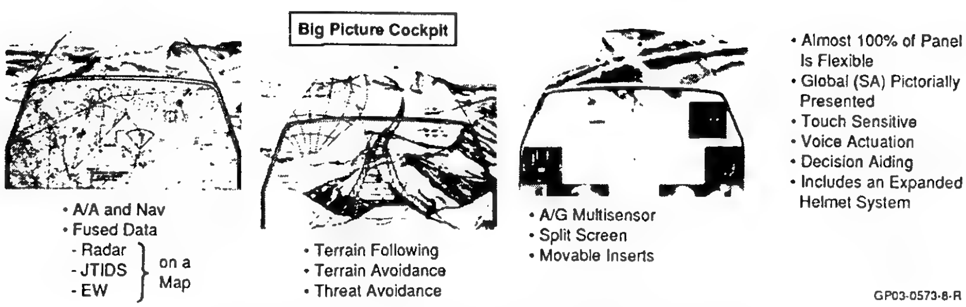


Figure 7. "Big Picture" Provides Total Reconfiguration Flexibility

THE FINAL FRONTIER

The laser, CBR and high energy weapon threat will require radical approaches to protecting the crew and providing sufficient information to fly and fight. There are two broad alternative solutions: 1) Remove the crew from the cockpit and fly and fight using remotely controlled vehicles. 2) Protect the crew within a "windowless cockpit".

Remotely Piloted Vehicles - Two technologies are necessary to provide this capability: 1) Sensors equivalent to the eye/brain are required to capture the visual combat scene real-time. 2) A secure, wide-bandwidth data link is required with near real-time capability to allow a pilot to fly and fight from a remote location.

For convenience, we will not treat this case because, SAM's, cruise missiles and other weapons fill many of these mission functions and the technology and frequency spectrum required for the immense amount of data to be linked between the pilot and vehicle on a real-time non line-of-sight basis make it impractical for any large number of fighters.

Windowless Cockpit - Needless to say, the concept of a sleek fighter without a canopy will cause most pilots to shudder and gag. However, the laser threat is real, they are in the field and 50 mile, zero time-of-flight "dazzles" are on the horizon.

For simplicity let us assume that sensors can provide spherical coverage around the aircraft with visual acuity. With the windowless cockpit concept there are two broad solutions: 1) Retractable protection whereby the pilot flies visual or non-visual depending on the situation and trains both ways. 2) Full time, enclosed cockpits with no outside vision. Both solutions require helmet displays and fixed displays, however, the retractable protection scheme has the disadvantage of having to meet 1000 times the ambient brightness requirement of the fully enclosed alternative.

Helmet vs Cockpit Displays - Without enormous breakthroughs in optics and display devices, the goal of a helmet display that does everything and doesn't require additional head-down displays does not seem practical for the high g environment in the near term. As shown in Figure 8, Cockpit and Helmet Displays are complementary. Both are required and both need extensive R and D to meet the needs of all three generations of cockpits discussed herein.

SUPPORTING TECHNOLOGIES

A number of supporting technologies are needed to gain the full advantages of the three generations of cockpits proposed herein. The real issue, however, is the cost/benefit ratio of individual and combined technologies. These are difficult questions to answer definitively because simulations and tests tend to emphasize environments whereby tested technologies are useful when nobody knows what the eventual distribution of scenarios will actually be. Fortunately, the aerospace industry and D.O.D. have seasoned design teams that are very good at getting the right systems in the final version of new generation aircraft.

Cockpit Displays

- Global Situation
- Plan View
- Beyond Visual Combat

Helmet Display

- Tactical Situation
- Perspective View
- Visual Combat

GP03-0573-7-R

Figure 8. Cockpit and Helmet Displays, They Complement Each Other - Both Are Required for SA

SUMMARY

The HUD and Multifunction Display cockpit using 5" and 6" CRTs have served us well. They have, however, two weaknesses: 1) No off-axis designation and information but this can be solved with Integrated Helmet Systems, and 2) No fused sensor and NAV data to a common range and coordinate system. This solution requires a large display, which most likely will be a non CRT technology.

The 90's will see a juncture of technologies such as flat panels, speech, graphics, decision aids, and immense computational capability ripening for the cockpit designers picking. Mission and vehicle requirements will and should drive the final choices.

Figure 9 shows the evolution of cockpits in the 20th century.

Eugene C. Adam Tel: (314) 234-1060
MDC Senior Fellow Fax: (314) 233-4433
Advanced Crew Stations
McDonnell Aircraft Co.
Mail Code 064 2203
P.O. Box 516
St. Louis, MO 63166

Mr. Adam is an electronics engineer and has over 35 years of aerospace cockpit design experience including the F-3H, F-101, F-4, F-15, F-18 and numerous other advanced aircraft. The F/A-18 Hornet has set the standard for "glass cockpits" currently in production. Mr. Adam served as a Navy aircrewman and has numerous patents to his credit. He was selected as one of the first MDC Fellows and is an internationally recognized innovator in cockpit design.

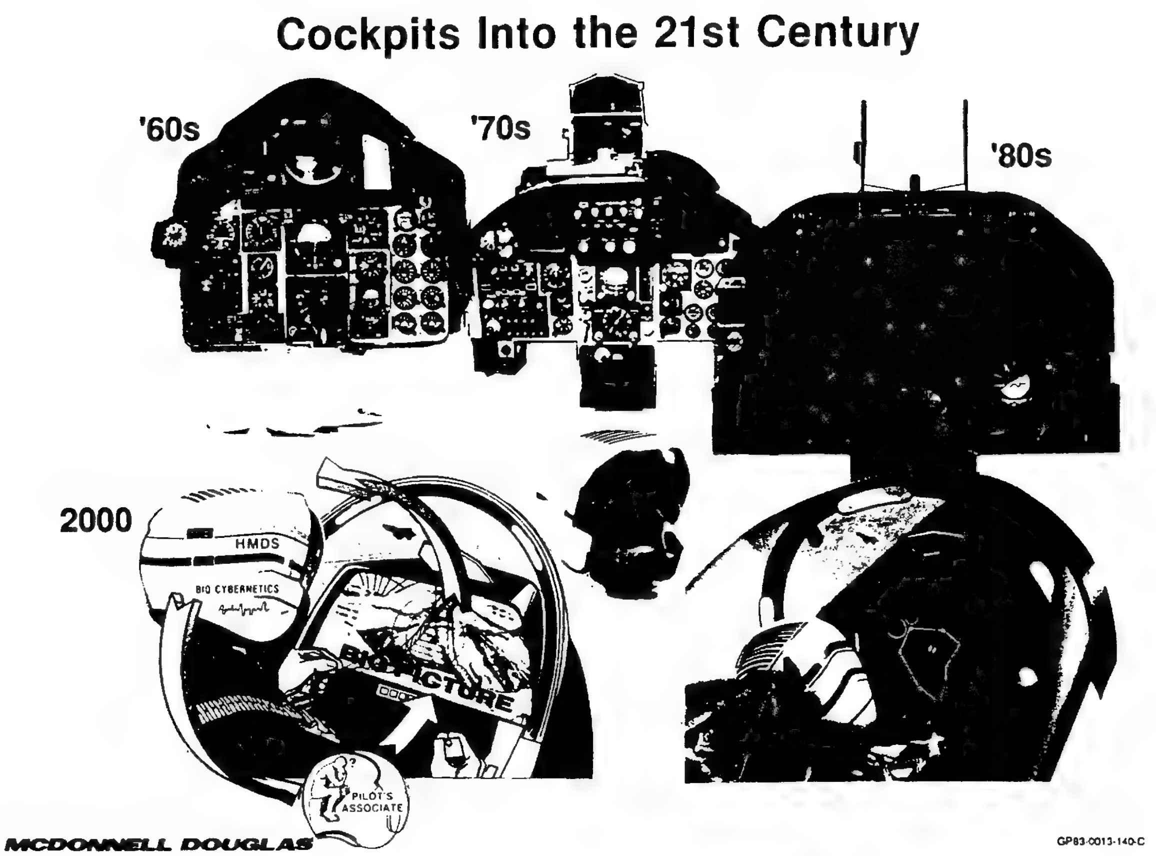


Figure 9. Cockpit Evolution in the 20th Century

DEVELOPMENT AND EVALUATION OF THE AH-1W SUPERCOCKPIT

D.J. Puleston
Principal Systems Engineer
GEC-Marconi Avionics Ltd.,
Airport Works, Rochester, Kent, England ME1 2XX

C.D. Holley
Principal Engineer
Bell Helicopter Textron Inc., Fort Worth, Texas 76101
USA

SUMMARY
An extensively modified AH-1W SuperCobra has been proposed for the British Army's new attack helicopter. Called Venom, the aircraft features an advanced technology mission equipment package integrated at the human/machine level by the AH-1W SuperCockpit™. This cockpit provides one of the world's most capable and integrated attack helicopter crew-stations, incorporating the latest techniques in "glass cockpit" design with key design objectives being to reduce crew workload, enhance mission effectiveness and maximise situational awareness.

INTRODUCTION
GEC-Marconi Avionics Ltd. and Bell Helicopter Textron Inc. have teamed to offer the Venom aircraft for the UK Army Attack Helicopter programme. The Venom is based on the Bell AH-1W SuperCobra aircraft, which has an excellent record of reliability and maintainability and is marinised for use by the US Marine Corps, who have operated this aircraft with distinction. The Venom programme takes the AH-1W airframe and refits it with the latest technology avionics, giving the aircraft a huge increase in overall mission capability and operational effectiveness. A joint development programme by the two companies has produced the SuperCockpit™, which is a low-risk upgrade for both new and retrofit aircraft.

SUPERCOCKPIT

Avionic and Aircraft Integration
The SuperCockpit™ is the heart of the approach to giving the Venom a huge increase in mission capability relative to the AH-1W, whilst reducing the workload associated with operating the aircraft and its systems. The Bell OH-58D aircraft has already demonstrated the benefits of applying technology to the low-level military helicopter and the Venom takes this approach still further. As a result the user will be able to maintain greater situational awareness and have increased effectiveness in the battle-field. The approach to integration has been to use extensive automation where judicious, to target high work-load drivers, to make the right data available at the right time and to achieve a level of effectiveness for the weapon system which was greater than the "sum of its parts".

Cockpit
The SuperCockpit™ provides two tandem crew-stations and is fully compatible with the existing AH-1W structure such that its inclusion in new aircraft (such as the Venom) or for retrofit is relatively straightforward. Preferred pilot crew-station is now the front seat, although the aircraft has almost identical functions/layouts in both crew-stations (Figs. 1 and 2). Crew vision has been emphasised and achieves approximately a 20% improvement over the existing AH-1W. The cockpit geometry has been reworked for improved anthropometric accommodation and ease of use, for example with a control/display interface within zone 1 reach.

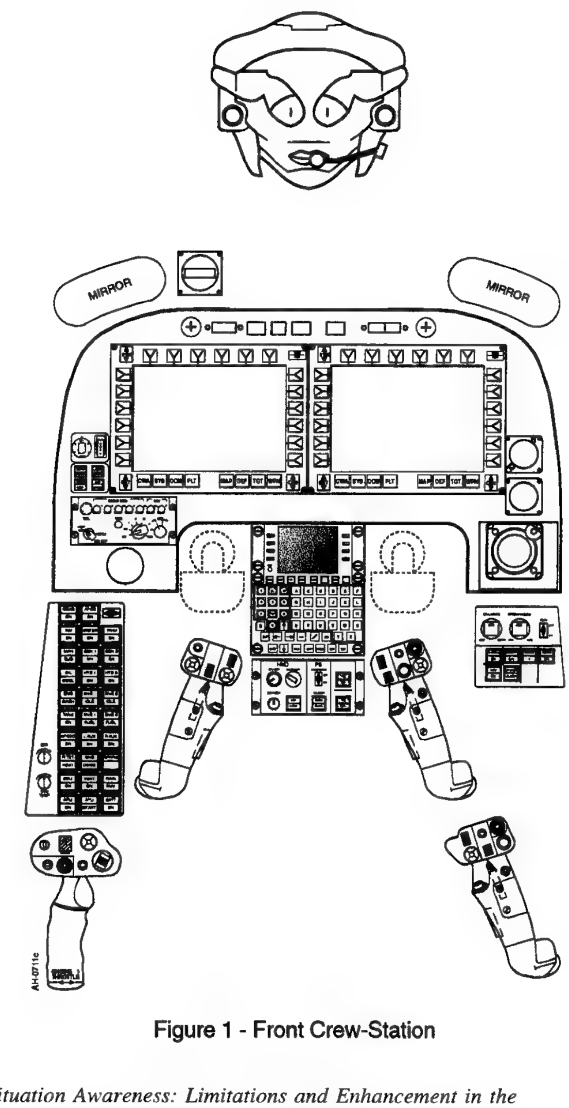


Figure 1 - Front Crew-Station

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

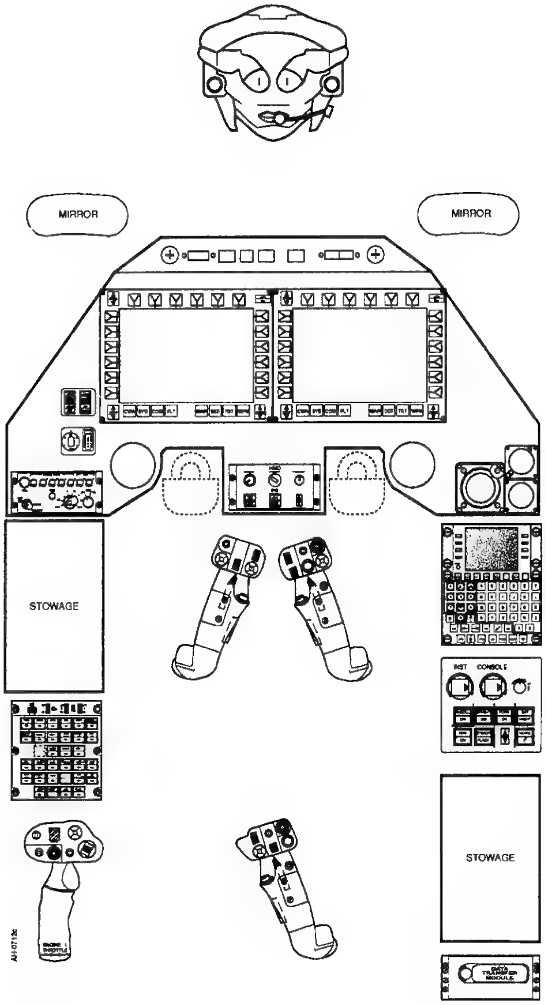


Figure 2 - Rear Crew-Station

Glass Cockpit Technology

Multi-Function Displays (MFDs) have been used in order to provide control and display functions for an ever increasing number of systems, with a number of current aircraft already reaching saturation point in terms of the number of controls, dials etc and the workload associated with using them. From a human factors engineering perspective, glass cockpits present a tremendous challenge for designing a system that deals effectively with managing the vast amount of information potentially available for display via the onboard computers and sensors. Having all this data available does not facilitate mission success and safety unless the crew has easy access to the correct information in a timely manner (Ref. 1).

The approach to the Venom display formatting has ensured that all display pages are available within two key presses (excluding ground crew maintenance pages etc) and that the most commonly sought information is available within one key press. Both crew-stations include two identical colour high resolution displays with hard keys (Fig. 3) to access information using a logically intuitive separation of functions. The theory behind the display design is scheduled for publication (Ref. 2). Presentation of information is logically consistent across all display surfaces including HMD.

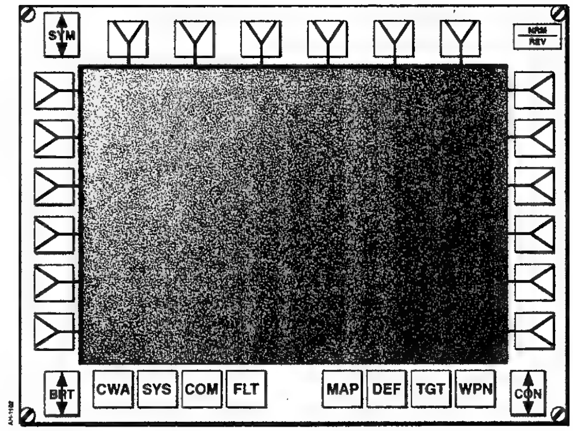


Figure 3- MFD Bezel

Head-Out Operation

In order to maximise crew head-out time for both piloting and situational awareness, the Venom has been equipped with two Helmet Mounted Displays (HMDs) from GEC-Marconi Avionics. These provide the crew with the ability to see symbology eyes-out either alone or combined with sensor/image intensifier (I^2) imagery. The HMDs (Fig. 4) are fully integrated and add functionality to the crew's helmets without adversely affecting weight or balance. A binocular approach has been used to avoid problems with binocular rivalry.

Figure 4 - Integrated Helmet Mounted Display

As the time to switch from head-up eyes-out operation to head-down eyes-in operation can be significant (Menu (Ref 3) has suggested that this time could be as much as 700 ms), a full complement of flight symbology has been provided, together with appropriate defensive, offensive and navigation functions (Fig. 5). The approach to helmet symbology has been derived as a result of three years investigation at the companies' simulator facilities and flight trials experience. Extensive use of real-world conformal symbology has been made, with the "fixed-wing" flight path marker being a popular addition. Research by Haworth and Seery has also supported the use of world referenced symbology (Ref. 4).

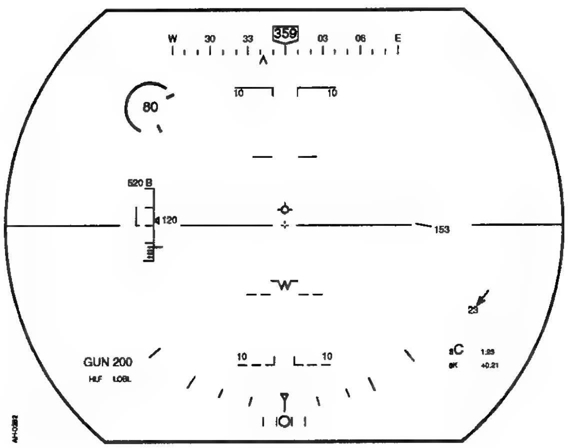


Figure 5 - Helmet Symbology

Night Operation
Extensive night operation capabilities have been added to the aircraft including a head steerable infra-red piloting sensor and a targeting sensor. These sensors are augmented by image intensifiers integrated within the HMDs and the whole cockpit is therefore fully NVIS compatible.

Battlefield Operations
The entire offensive and defensive capabilities of the aircraft have been revised and significantly improved. A steerable targeting sensor using a second generation infra-red detector provides high resolution sensor imagery and the ability to detect and identify targets at very long ranges. This imagery is available on the MFDs, with the copilot/gunner no longer being forced head-down into a fixed sight with the associated loss of situational awareness through not being able to look outside or even inside the cockpit. Displaying the targeting imagery on the MFDs has the secondary benefit of making it available in both crew-stations, potentially offering a second opinion on the "enemy" nature of the target, although identification is in any event significantly improved by the unrivalled image quality of the Venom targeting sensor. In developing an MFD based system, extreme care was given in ensuring that the installation, resolution, synchronisation and eye characteristics were properly considered so as to meet stringent range performance requirements .

In addition to range performance, the Venom has the capability to track multiple targets simultaneously and to control multiple missiles in the air. Target/friendly positional information is readily transmitted by digital burst communications to all other aircraft on the mission for transfer of targets or for general situational awareness. The Venom also has the capability to perform a quick scan of an area for subsequent replay when masked. As a result, the mission commander could unmask briefly, perform a scan, remask, locate and identify targets from the stored scan, and transmit the information to the other aircraft. Using the Venom "Brimstone" fire-and-forget missile, the other aircraft could then fire multiple missiles from cover and need never be exposed to the targets.

Accurate Navigation and Positional Awareness
Navigation has always been a major task for the copilot of an attack helicopter and it has been shown that over ninety percent of the copilot's time can be allocated to the navigation task (Ref. 5). Each crew-member is therefore provided with a digital map, which shows present position as derived from an accurate IN/GPS system. Research has shown that this type of equipment reduces: the time devoted to navigating, inter-crew verbal messages, visual workload, navigation errors and deviation from track (Ref. 6). The map display may be overlaid by various overlays including: route, intelligence information, other aircraft and threats (Fig. 6).

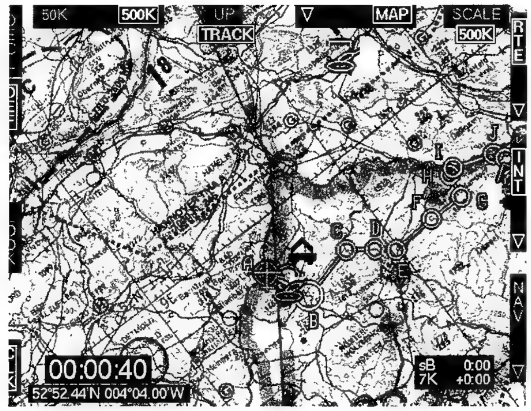


Figure 6 - Map Display

Hands-On Control
The forward cockpit's cyclic control stick is located on the right console. The cyclic grip contains ten switches to provide hands-on control of the following functions: radio/intercom transmit, force trim, weapons select, SCAS disengage, HMD video (IR/I^2), weapons action/steer, missile cage/uncage, HMD declutter, display select, and weapons fire (trigger). The aft cockpit uses the same grip mounted on a kneeling, centre-stick. The kneeled position provides clearance for using mission grips when these are unstowed.

The collective control stick is located on the left side of the crew station and contains two twist-grip throttles for engine power management. The forward throttle is integrated with the collective grip and provides enhanced tactile response relative to the existing AH-1W. The collective grip contains seven switches to provide the operator with hands-on control of the following functions: radio frequency select, idle stop release, emergency jettison, countermeasures, searchlight control, searchlight slew, and hover hold.

Each cockpit contains two stowable mission grips, located below the MFDs, that are installed on telescoping platforms. The grips pivot and rotate to an upright orientation when moved from the stowed to the operational position. In addition, the telescoping mount provides lock-type positions for accommodating fore and aft adjustment. The left mission grip has nine switches for controlling the following functions: TV/IR focus, TV/IR gain and level control, laser fire (trigger), LOS acquire, FLIR polarity, track box size adjust, sensor select, action steer, and FOV select. The right mission grip has ten switches for controlling the following functions: weapons fire (trigger), weapons select, turret/cursor slew, track function select, gun targeting select, IR auto initiate/manual, HMD video, HMD declutter, missile cage/uncage, and weapons action/steer.

EVALUATION
The current SuperCockpit™ configuration has been supported by a strong commitment to man-in-the-loop simulation at both GEC-Marconi Avionics and Bell Helicopter Textron. The design team has been grateful to

the many pilots from the UK Army Air Corps, US Marine Corps, test pilot schools and government research organisations who have contributed to the design with helpful feedback and suggestions.

A first stage prototype cockpit was formally evaluated in Bell's full mission simulator during November 1992 using six pilots with AH-1W and military experience (four active duty USMC pilots and two Bell test pilots), divided into three crews. Training and evaluation was conducted at the crew level, with questionnaires completed at the individual pilot level. Each crew received 12 hours of ground school and three hours of actual flight training coupled with 12 hours of vicarious flight training. Following training, each crew flew a series of evaluation scenarios that totalled 2.5 hours and included missiles, rockets and guns. Training and evaluation time was equally divided between front and rear cockpits for each pilot. An extensive questionnaire, using Likert-type ratings supplemented with qualitative comments, was used as the data collection instrument. The pilots were also asked to list the three best and three worst design features. The three best were digital map (overwhelmingly so), pilot in front seat, and HMD. The three worst were collective grip design, mission grip location, and location differences of the targeting system transducer switch. Each of these evaluative comments was considered, along with specified and mission-derived customer requirements, in the evolution of the SuperCockpit™ in its current Venom configuration.

The simulators at both companies are being upgraded to the latest configuration, including multiple target tracking and multiple simultaneous missiles in the air. The ability to use the simulators for rapid prototyping in the design has unquestionably assisted and improved the design process and indeed simulator, CAD equipment and word processing facilities were all networked at the outset for the generation of the original Venom control and display specification.

By combining the proven reliability, marinisation and maintainability of the AH-1W SuperCobra airframe with the latest high technology and integrating this in the SuperCockpit™, GEC-Marconi Avionics and Bell Helicopter Textron have designed the highly capable Venom helicopter, which is capable of fulfilling a wide range of missions in day/night/poor weather. The Venom offers a high capability, low cost solution to customer requirements well into the next century - both in the UK and world-wide.

REFERENCES

1 Graf, V.A., and Holley, C.D., "Human Factors Impact on the V-22 Osprey Cockpit: An Overview", American Helicopter Society, 44th Annual Forum, Washington, D.C., 1988.

2 Holley, C.D., and Busbridge, M.L., "Evolution of the Venom Variant of the AH-1W SuperCockpit", American Helicopter Society, 51st Annual Forum, Fort Worth, Texas, 1995.

3 Menu, J.P.R., "Head-Up/Head-Down Transitions: Measurement of Transition Times", Aviation, Space and Environmental Medicine, March 1986, pp218-222.

4 Haworth, L.A., and Seery, R.E., "Helmet Mounted Display Symbology Integration Research", American Helicopter Society, 48th Annual Forum, Washington, D.C., 1992.

5 Sanders, M.G., Simmons, R.R., and Hofmann, M.A., "Visual Workload of the Copilot/Navigator during Terrain Flight", Human Factors, 1979, 21(3) pp369-383.

6 Cote, D.O., Krueger, G.P., and Simmons, R.R., "Helicopter Copilot Workload during Nap-of-the-Earth Flight", Proceedings of the Second Symposium on Aviation Psychology, pp152-162.

POINT-OF-GAZE MEASUREMENT IN AVIATION RESEARCH

by

H.A. Mooij
9 Leidsestraatweg
2341 GR Oegstgeest, Netherlands
Phone: +31(0)71-170201
Fax: +31(0)71-174621
E-mail: mooij@pop.knoware.nl

1 INTRODUCTION

Traditional assessment methods (performance and subjective) may not be sufficient for the evaluation of man-machine systems, in which an appropriate level of situation awareness of the operator is of crucial importance. Eye Point-Of-Gaze (POG) and eye-blink measurement is one of the psychophysiological methods which may be helpful, as it can be continuously available without being intrusive to the operator's task.

Eye point of gaze can provide data about where in the environment information is sought, as well as about the pattern of eye-scanning as evidenced in different situations.

After a feasibility study, subsidised by the Netherlands' Agency for Aerospace Programmes (NIVR) in 1990, the European Space Agency (ESA) contracted Mooij & Associates in mid-1991 to develop a system capable of determining point of gaze in real time in digital form, for the evaluation of competing designs of user interfaces for controlling life-support systems as well as scientific experiments on board future space craft. The system was developed over a period of three years. A successful pilot experiment on a Graphical User Interface, also performed under ESA contract, concluded the initial development of the system at the end of 1993.

From the beginning of 1994 onwards, the system - the commercial version of which is called OBSERVER - underwent many significant improvements mainly in the areas of accuracy and user-friendliness (Ref. 1, 2).

The application of eye point-of-gaze data in aviation research will be set out below (Chapter 2), followed by a description of the OBSERVER system in Chapter 3. In conclusion, an overview of OBSERVER usage in two simulator programmes is given in Chapter 4.

2 APPLICATION OF POG INFORMATION

2.1 General

New information technology for the enhancement of situation awareness is being introduced in semi-automatic man-machine systems. The challenge now is to design the man-machine interface and the automation structure in such a way that the situation awareness aimed for is actually obtained. The following one-sentence definition of situation awareness presented in an informative article on the subject, (Ref. 3) is adopted:

Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and projection of their status in the near future.

When new information technology is used to enhance the situation awareness of system operators, the man-machine interface layout and the cognitive workload are central issues to be regarded during the design and development phases.

Research and development evaluations have been reported in which eye trackers have been used in the analysis of perceptual motor tasks, such as driving a car and flying an aircraft. In most cases, the data was obtained from the recorded images of

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

a "scene camera" overlaid with a cursor indicating the measured eye line of gaze. Frame-by-frame digitising is required after the tests in order to enter point-of-gaze information into a computer for the purpose of analysis. When the emphasis lies on tasks which are highly "cognitive", an operator in many cases gets information from displays on fixed screens (e.g. simulators). For these cases, OBSERVER with its automated point-of-gaze determination in real time, delivers data for immediate or even on-line analysis. Some application areas for point-of-gaze data will be introduced below.

2.2 Man-Machine Interface Development for Manned Spacecraft

As more computer support and control are introduced, the operation of systems by crew on board spacecraft is changing drastically. There are very few flight opportunities, which means there is also little opportunity for evolutionary development of new systems or building up confidence through regular use.

In the International Space Station, the crew will have a variety of tasks in both spacecraft system control and on-board experiment control. It is known that crew time, on board as well as on the ground (for training), will be limited, resulting in conflicting requirements: the crew being involved in ever more activities while having ever less time available for training on each particular function.

The need for a tool with which to thoroughly evaluate competing designs of graphical user interfaces for controlling both spacecraft systems and on-board experiments, formed the motivation for the development of the OBSERVER system.

2.3 Aviation Research

The following are examples of experiments in which eye point-of-gaze data have been used and reported in the open literature:

Head-Up Display Symbology/Mooij & Associates (Ref. 4)

The purpose of the experiment was to explore the feasibility of using OBSERVER when it comes to recording and analysing data about a pilot's point of gaze on the Head-Up Display (HUD) when flying a fast-jet aircraft. This was a pilot experiment for studies currently conducted by the National Aerospace Laboratory, NLR. The experiment is discussed in some more detail in Chapter 4.

Head-Down Display Configuration/Airbus-Aeroformation (Ref. 5)

Airbus/Aeroformation performed an experiment to determine whether there was a difference in search strategy and acquisition of information by pilots between Airbus A310 - with two Cathode Ray Tubes (CRT) (on top of each other) plus classical electromechanical flight instruments - and Airbus A320, with two CRTs (side by side) which included all primary flight control information. In this experiment, the eye point of gaze of both the Captain and the First Officer were measured.

ATC Data-Link Message Exchange/NASA (Ref. 6), Berlin U. (Ref. 7), Boeing plus NASA (Ref. 8)

Air Traffic Control (ATC) message-exchange between aircraft and air-traffic control ground facilities via digital data link raises intriguing questions regarding a division of attention which have to be answered in the near future. In the above references, various aspects are studied. Elements of the studies are:

- Pilot and co-pilot scanning behaviour, comparisons of data-link protocols with the conventional voice radio approach (Ref. 6). In this experiment, the eye point of gaze of both the Captain and the First Officer were measured.

- Investigation of the feasibility of visual display of ATC messages (data link) in advanced glass cockpits, in particular in the Navigation Display of an Electronic Flight Instrument System (EFIS) (Ref. 7).

- Study of the influence of data link on scanning activities - of both the environment and the instruments - impact of systems integrated into the Flight Management System (FMS) versus typical retro-fit system implementation with a separate interface device (Ref. 8).

ATC Controller Working position/NLR (Ref. 9)

A consortium of parties is working on a detailed specification for the future (air traffic) Controller Working Position (CWP), the definition of the CWP characteristics and the assessment of a suitable man-machine interface. An eye point-of-gaze experiment, performed in the same framework is introduced in Chapter 4.

ATC Decision-Aiding System/NLR (Ref. 10)

CTAS (Center-Tracon Automation System) work is currently being conducted at NLR. The purpose of this study is to compare three levels of possible future ATC automation, using a prototype decision-aiding system. The three levels are: "traffic information level", "conflict information level" and "solution automation level". The study, in which the OBSERVER system is used, should lead to the selection of one of these levels for future implementation.

2.4 Advanced Applications

Point-of-Gaze Data for System Control

In searching for new and better interfaces between systems and their users, it can be very useful to exploit an additional mode of communication between the two parties. Typical human-computer dialogues are rather biased in the direction of communication **from** the computer **to** the user. Animated graphical displays, for example, can rapidly communicate large quantities of data, but the inverse communication channel has a very low bandwidth. The availability of an additional, rapid information channel **from** the user **to** the computer would be helpful, particularly if it requires little effort on the part of the user.

Using point-of-gaze data to drive a (computer) system can be desirable, because:

- Point of gaze has a high bandwidth due to the fact that eye muscles, being extremely fast, are able to respond more quickly than most other muscles.

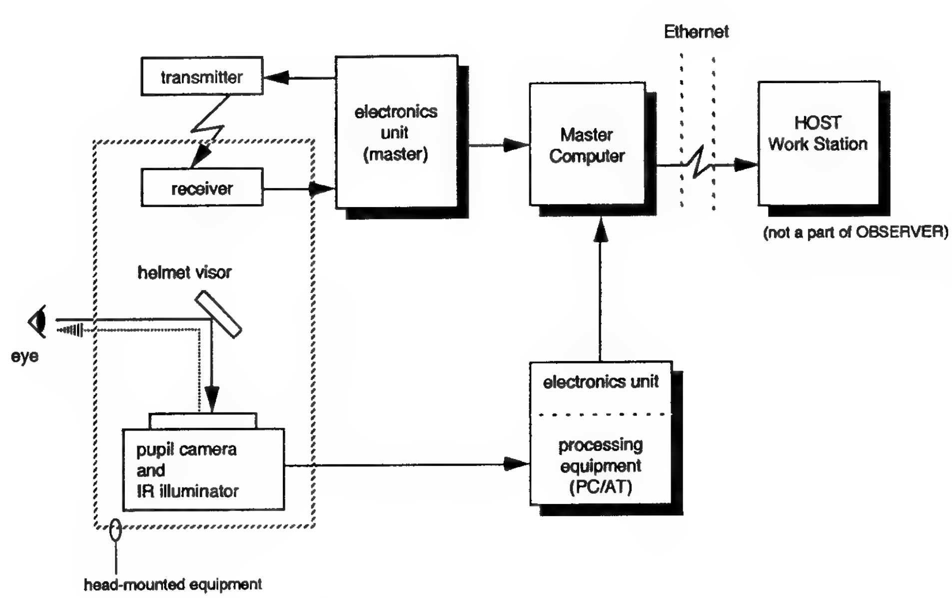
- Point of gaze, based primarily on eye motion, can be beneficial under high-g loading (eye motion under high-g loading is perfectly feasible).

- Shifting point of gaze comes naturally and requires no conscious effort.

The arguments mentioned above demonstrate that point of gaze is a potentially useful, additional user-computer input channel, especially in situations where the user is already heavily burdened.

Since eyes continually dart from point to point in rapid and sudden saccades, unfiltered point of gaze cannot simply be used to replace computer input devices such as the mouse. This is why point-of-gaze fixations should be used. System control

Fig. 1



OBSERVER, POINT-OF-GAZE MEASURING SYSTEM

using point-of-gaze fixation data from OBSERVER has already been demonstrated.

Bi-lateral exchange of information

There is a growing interest in "interactive ergonomics" which addresses the various disciplines of cognitive sciences based on the concept of parallel coupling of man and machine with a two-way exchange of information (Ref. 11). Non-intrusive psychophysiological measures, such as eye point of gaze, may very well become a permanent part of certain man-machine systems. The machine monitors the selected psycho-physiological parameters of an operator and issues messages accordingly.

3 DESCRIPTION OF THE OBSERVER SYSTEM

3.1 System

The OBSERVER system is capable of providing point-of-gaze data in real time. This characteristic makes it possible to use the system as a high-bandwidth designation tool (information **from** the user **to** the computer). Figure 1 depicts the OBSERVER system in the form of a block diagram.

During the design of the system, special attention was devoted to ensuring user-friendly calibration features. A description of the subsystems constituting OBSERVER, the output and performance is given below.

3.2 Subsystems

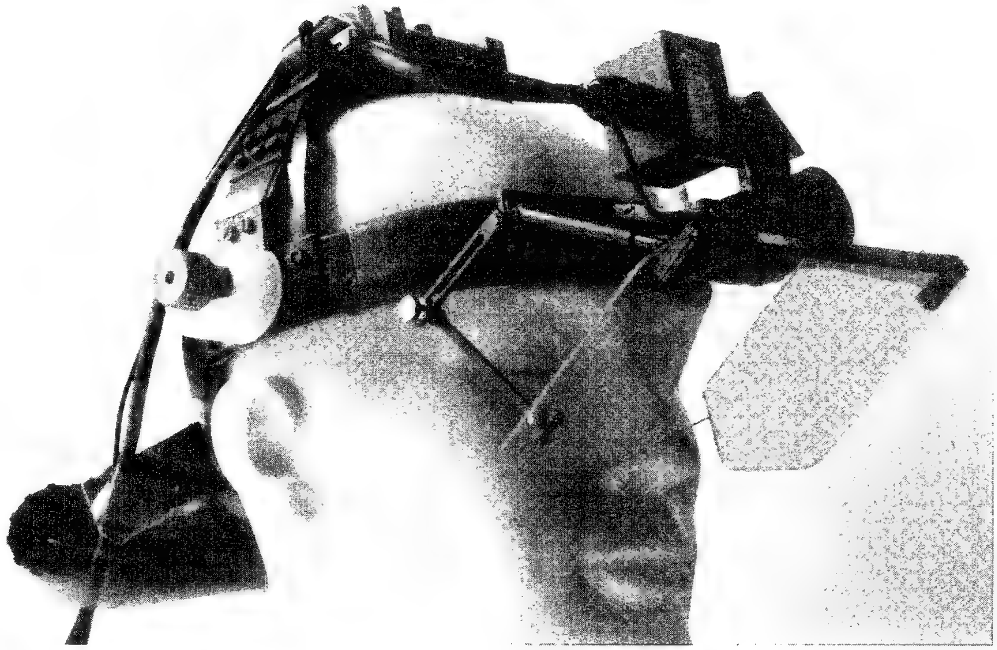
OBSERVER consists of three major subsystems: the Eye-tracking subsystem, the Motion-tracking subsystem and the Calibration/preprocessing subsystem.

Eye-tracking subsystem

The Series 4000 Eye Tracker with head-mounted optics of Applied Science Laboratories (ASL) forms the basis of the eye-tracking subsystem. The mass and inertia of the head-mounted optics, which has no peripheral vision restrictions, is of a level allowing prolonged wear. Figure 2 shows a photograph of the head-mounted optics.

The eye-line-of-gaze tracking range is 50(H)x40(V) degrees, with an update rate of 50 samples/second. Eye-calibration time is short, while the accuracy is 1 deg (rms).

Fig. 2



HEAD-MOUNTED OPTICS

MOOIJ & ASSOCIATES 54-05

The technique used for eye-tracking is the pupil-to-corneal reflex vector method of the "bright pupil" variety.

The subsystem is controlled through a Control Unit and a dedicated 486 PC/Monitor combination. An "eye monitor" is used during the tuning of the head-mounted optics.

To calibrate the eye tracker for a particular person, a short routine is performed during which data are loaded while the person alternately looks at nine different points on a calibration card, which is either temporarily fixed to the head or part of a head clamp.

Motion-tracking subsystem

The "magnetic type" position- and orientation-measuring system of Ascension Technology Corporation (Flock Of Birds), indicated here as motion-tracking subsystem, consists of a transmitter and a receiver both attached through cables to an Electronics Unit. The transmitter is the fixed reference against which the receiver measurements are made, while the receiver is attached to the headband also holding the head-mounted optics. The system works on the basis of a pulsed DC magnetic field. "Mapping" of the environment is not required. The position and orientation of the receiver is measured anywhere within a sphere of 0.9 m radius, with an accuracy of 0.3 cm rms for the position and 0.5 deg rms for the orientation. The system has a maximum update rate of 100 samples/second.

To be able to determine the vector describing the receiver-to-eye separation, a short calibration routine is executed by the subject under guidance of a test director. A stylus and an optical sight are the tools for this part of the calibration.

Calibration/preprocessing subsystem

The calibration/preprocessing subsystem consists of a "master" computer/monitor (Apple Macintosh) and a computer programme named EPOG. The programme incorporates three driver modules for communication with the eye-tracking subsystem, the motion-tracking subsystem and the network driver (Ethernet). The function of the calibration subsystem is fourfold:

- It provides a means to enter data during calibration. In this phase, the programme determines the position of the subject's eye in the system of coordinates of the (head-mounted) magnetic receiver.
- By using the position- and orientation measuring system, it provides a means to enter the position (in a room-fixed reference system) of the multiple surfaces to be observed by the subject.
- It performs all actions required with respect to calibrating the system and measuring/-recording point-of-gaze fixations. Measuring and recording point-of-gaze fixations may be remotely controlled from another computer system connected with OBSERVER via Ethernet.
- It facilitates the selection of certain parameters in the software, e.g. temporal and angular thresholds in the calculation of point-of-gaze fixations.

All four functions mentioned above are selected and controlled by means of a Dynamical Graphical User Interface (DGUI). All commands and selections etc. are given through a mouse (or tracker ball). Only the names of files and system settings are entered through the keyboard.

Provided there is adequate memory space in the Apple Macintosh computer as well as a statistical analysis programme, no additional computing facilities are required. When OBSERVER is used in simulator experiments, data should be exchanged with the process in the simulator computer. To facilitate real-time data exchange via Ethernet, the EPOGClient software package is available as part of OBSERVER.

3.3 Output and Performance

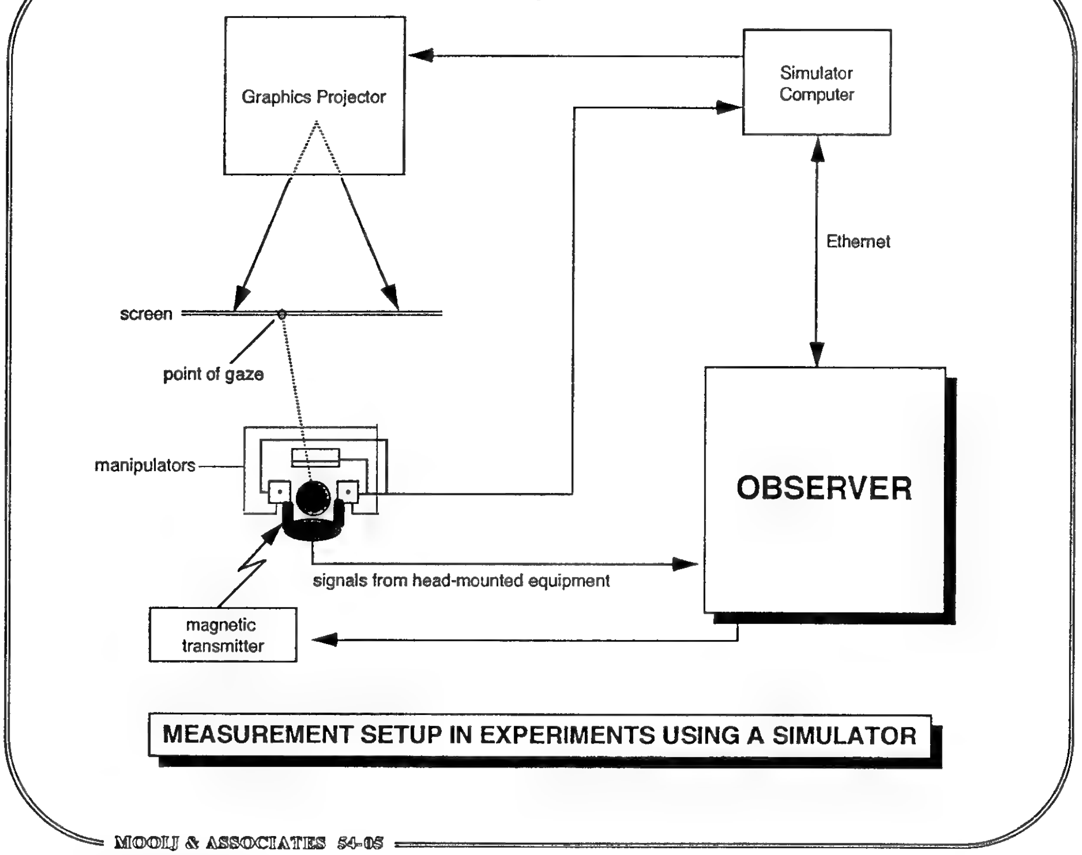
The preprocessed data related to point-of-gaze fixations are:

- Starting time of fixation
- X and Y of fixation
- Duration of fixation
- Pupil diameter
- Surface identification
- Distance eye to surface.

The update rate of real-time point-of-gaze is 50 Hz.

For setting up or monitoring whilst the system is in "recording mode", the intersection of a pair of cross hairs on the monitor of the master computer represents point of gaze (updated 50 times per second) and a fixation trail represents the most recent fixations recorded.

Fig. 3



4 TWO EXPERIMENTS WITH THE OBSERVER SYSTEM

In literature on visual perception, temporal-spatial patterning and the duration of fixations are regarded as a reflection of the perceptual strategy used by an observer to extract meaningful information from a display.

The duration of a fixation period most likely implies the relative importance of the display area to the observer and is often interpreted by researchers as a measure of covert cognitive processing.

An accurate synchronisation of the measuring device delivering point-of-gaze data and the controlled process (e.g. vehicle control) is mandatory in the support of detailed post-experimental analysis and interpretation of recorded point-of-gaze data. Figure 3 presents a possible solution for the application of the OBSERVER system, in case the controlled process resides in a simulator computer. Although a graphics projector/screen combination is depicted in the figure, any form of simulator visual display may be used.

Two experiments are described below in which the OBSERVER system was applied in studies related to man-machine interfaces in aviation.

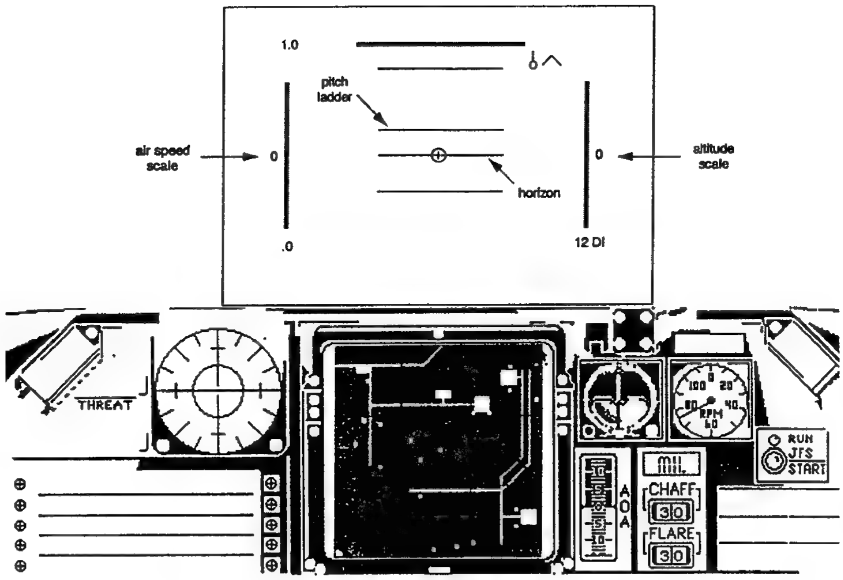
HUD Scanning

Upgrading of the avionics suite of existing fighter aircraft (F-16) poses questions to the R&D community regarding the optimum quality of the pilot/system interface. To improve insight into the usage of visually presented information in the cockpit, the National Aerospace Laboratory NLR is in the process of conducting a series of research projects.

Under contract of the National Aerospace Laboratory NLR, Mooij & Associates evaluated the combined use of the OBSERVER system, a PC-based flight simulator and an off-line computer programme for statistical analysis of measured point-of-gaze data.

The objective of the experiment was to explore the possibilities of the OBSERVER system when it comes to recording and analysing a pilot's point of gaze in and outside the cockpit, when flying a fast-jet simulator. In addition, it was meant to provide

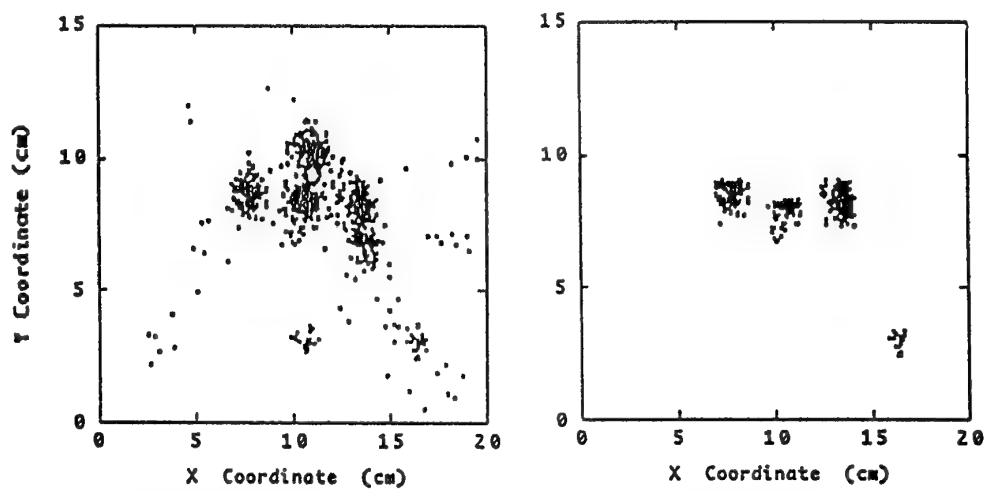
Fig. 4



FALCON MC FORWARD VIEW IN COCKPIT

MOOIJ & ASSOCIATES 54-05

Fig. 5



UNFILTERED FIXATIONS
SECOND FLIGHT PROFILE

FILTERED FIXATIONS
SECOND FLIGHT PROFILE

FIXATIONS DURING ONE FLIGHT PROFILE

MOOIJ & ASSOCIATES 54-05

an understanding of the characteristics of the data set obtained.
The subject's task was to perform a flight profile of 12 minutes' duration as precisely as possible, using a combat flight simulator (FALCON MC). Figure 4 shows the forward view in the cockpit as presented by the simulator. Results were obtained from analysing the data of three identical sessions. Around 1000 fixations of 200 ms or more were recorded per session. Figure 5 gives the unfiltered and filtered fixations recorded during one flight profile. The filtered case shows the fixations related to speed, horizon and altitude indication on the HUD and RPM indicated on the instrument panel (lower right hand corner). The exploratory analysis resulted in statistical information per flight phase (a total of nine phases: "take-off", "turn", etc.). Examples of a few of the statistical indicators determined are: percentage of fixation time per variable (e.g. airspeed, altitude), transitions (number of fixation movements from one variable to another), interfixation times (time between successive fixations on the same variable).

Future Controller Working Position

The mental acquisition of data by the air-traffic controller is largely dependent upon the quality of the interface to and from the "machine". Typically, the controller is presented with dynamic displays intended to convey a picture of the current air traffic situation. This composite display features correlated radar and flight plan data, the route structure map and heavy weather indications. The controller inputs data in order to record decisions or to facilitate the capture of data.

In the EURET programme, the Commission of the European Union and the SWIFT consortium are working on the detailed specifications for the future Controller Working Position, on the definition of the Controller Working Position characteristics and on assessing a suitable Man-Machine Interface (MMI). SWIFT concentrates mainly on the quality aspects of the MMI, i.e. how do we present the information available in the flight data processing system in the most efficient and elegant way.

In March 1994, the National Aerospace Laboratory NLR used the OBSERVER system in combination with their air traffic simulator (NARSIM) to demonstrate that point-of-gaze information can be useful for the examination of screen layouts and of the way air-traffic controllers employ tables containing Area COnflict Detection (ACOD), Short Term Conflict Alert (STCA) and UPLINK information, (Ref. 9). Figure 6 shows the three screen layouts evaluated in the experiment. In the experiment, data link was the only mode of communication. Apart from a tracker ball, a Touch Input Device (TID) was used by the air-traffic controller

The data has shown that a very high percentage of the fixations (of 150 ms or more) were correlated with traffic on the screen (quite a few elements of information are presented with each aircraft symbol). In this particular setup, there were hardly any fixations on ACOD and STCA tables. Air-traffic controllers stated that the equipment used for point-of-gaze measurement was acceptable for their task in the setting presented to them. [At the beginning of 1995, the OBSERVER system was used again in a "CTAS study" using the same ATC simulator. The purpose of this study was to compare three levels of possible future ATC automation, using a research version of the decision-aiding system CTAS, Ref. 10.]

5 CONCLUSION

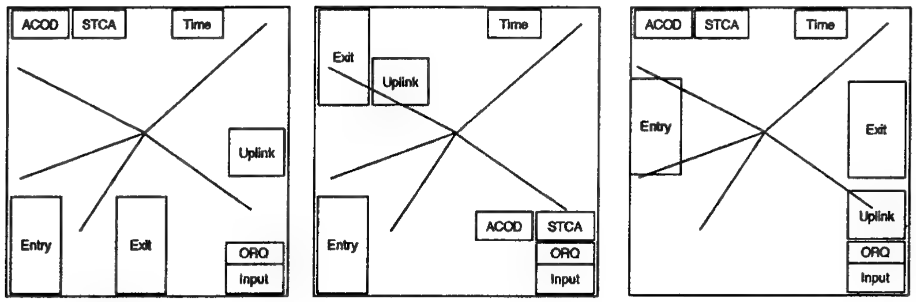
It is definitely a challenge to design man-machine interfaces and automation structures in such a way that situation awareness is maintained at an adequate level. The determination of non-intrusive psychophysiological measures, such as eye point of gaze, provides the basis for an accurate analysis of the scanning activities of operators forming a part of complex man-machine systems.

The OBSERVER system described in this paper has proven to be an efficient device for delivering point-of-gaze fixation data, without interfering with the person wearing the headband on which the electro-optical elements and a miniature magnetic receiver are mounted.

Programmes to explore the advantages of using real-time point-of-gaze data in advanced applications, such as system control and parallel coupling of man and machine, are in a phase of definition.

Fig. 6

(Display is 0.5 x 0.5 m)



Configuration A Configuration B Configuration C

ACOD = Area COnflict Detection STCA = Short Term Conflict Alert ORQ = OnReQuest

(the display presents the Dutch FIR route structure map)

"TABLE CONFIGURATIONS" DURING THE SWIFT EXPERIMENT

MOOIJ & ASSOCIATES 54-05

6 REFERENCES

1 Zon, G.D.R.; Mooij, H.A.; Bouwens, J., "A Non-Intrusive Way to Measure Point of Gaze", AGARD-CP-541, 1994.

2 Mooij, H.A.; Farkin, B., "OBSERVER for Point-Of-Gaze Determination in Real Time", Conference Proceedings of the Internation Training Equipment Conference and Exhibition, 1994.

3 "Situation Awareness: the Buzz Word of the 90s", CSERIAC Gateway, Vol. V: No. 1 (1994).

4 "The Use of OBSERVER for Analysing the HUD Scanning Behaviour of Fast-Jet Pilots", Mooij & Associates Report MA-94-05, 1994.

5 Racca, E.I., Instrument Scanning in New Cockpits", ICAS-94-8.7.1 and "Pilot's Psychology Through an Oculometric Study", Paper in the Proceedings of the EAAP Conference, 1994.

6 Waller, M.C.; Lohr, G.W., "A Piloted Simulation Study of DataLink ATC Message Exchange", NASA Technical Paper 2859, 1989.

7 Hüttig, G.; Hotes, A.; Tautz, A., "Human Factor Design Considerations for Air Traffic Control Information Displays in a Modern Glass Cockpit", Paper in the Proceedings of the EAAP Conference, 1994.

8 Infield, S.E. et al., "The Effects of ATC Data Link on Instrument and Enviroment Scanning During Flight Operations", Paper in the Proceedings of the EAAP Conference. 1994.

9 SWIFT Point-of-Gaze Experiment, Personal communication with Mr A. de Reus, April 1995.

10 CTAS Study, Personal communication with Mr B. Hilburn, April, 1995.

11 "Psychophysiological Assessment Methods", AGARD-AR-324, 1994.

SPEECH CHARACTERISTICS OF SITUATIONAL AWARENESS IN THE COURSE OF COPING WITH IN-FLIGHT EMERGENCIES

J. Šulc
Institute of Aviation Medicine
P.O.Box 19, 160 60 Prague 6
Czech Republic

SUMMARY

A pilot, posed with an in-flight emergency works often under considerable pressure of strong emotions, resulting from an imperative threat to life. They can either destroy the situational awareness or leave it unaffected. Sonographic analysis allows to distinguish the emotional strain from behavioural breakdown with sufficient exactitude. While the information on emotional activation is mediated through the pitch variations, its impact on behaviour can be deduced from the temporal course of utterances. A relatively reliable sign of the loss of situational awareness in life - threating situations is represented by the s.c. "open scissors phenomenon", formed by an antagonistic movement of the pitch in relation to the speech rate. Its essence resides in the uncontrollable effect of asthenic emotions, leading to the enhancement of muscular stiffness, which increases the pitch and retards the speech rate. External appearance of disadaptation to the emotionally demanding situation conceivably correlates with the impaired quality of perceptual and cognitive processes, forming a basis for the in-flight situational awareness.

1 INTRODUCTION

The training of pilots in managing the unexpected in-flight critical situations is an important part of a complex, life-long effort in reaching the aviator's mastery. The training process differs in some significant aspects from the acquisition of skills, forming the base of pilot's mission effectiveness. A great deal of considered in-flight emergencies can be exercised only in crucial points of the algorithm of their identification and management, what is undoubtedly the prerequisite of their successful solution. At the same time the majority of sudden impairments of regular missions, namely those with imminent threat to the life, cannot ever be trained with pertinent emotional component. Thus, despite the intensive simulator and cockpit resource management training of in-flight emergency procedures the pilot's real experience in coping with sudden non-standard situations remains limited. This accounts for some uncertainity in the prosperousness of pilot's performance, aimed to the solution of arising problems, especially when the flight came to a bad end and the interrogation of the crew is no more possible.

2 BACKGROUND

Human speech is capable of transferring a remarkably rich set of coplex and highly integrated information on various physical and psychological characteristics of the speaker. In the course of communication the human brain actively classifies and extracts information relevant to the needs and interests of the acting subject. In comparison with other instrumental methods used in the objectification of human response to the environmental stimuli, speech signal processing has certain

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

advantages. There is no need of special sensors, the acquisition as well as the processing of the records (as opposed to their interpretation) is relatively simple.

Only in 1954 von ESSEN (1) defined psychophonetics as an interdisciplinary scientific discipline, looking for the physiological basis and phonetic expression of mental processes, above all of the emontions. In aviation industry HECKER and collaborators (2), WILLIAMS and STEVENS (3,4) and members of the 1969 AGARD AMP Symposium were among the first, who recognised the potential benefit of electroacoustical analyses of air-to-ground communication for the assessment of task-induced stress. Russian researches in these times focused their interest on the voice representation of emotional strain in Soviet cosmonauts (5-8). Besides the relative graduation of emotional activation they could distinguish objectively the polarity of passing emotional states. NIWA (9,10) in Japan developed the s.c. vibration space analysis technique, demonstrating the experience of stress in urgent situations.

In the same time within the framework of a widely conceived research on aviation radiophony the acoustical and psycholinguistic pecularities of speech, produced in formidable in-flight situations have been explored in the Institute of Aviation Medicine, Prague (11-16). The main purpose of research activities was to help the Accident Investigation Board in clearing up the aviation accident or incidents with a not entirely plain course of events. Meanwhile the chance to process the peculiarities of verbal behaviour of the first Czechoslovak cosmonaut during space flight were exploited as well (17).

The results of many years' standing research contributed not only to the examination of the aviation mishaps, but also to the complementation of theoretical conceptions of radiophony as an important factor of a joint working activity of all air-traffic's and even of space flight's participators. Nevertheless a considerable amount of psychofhysiological analyses of the aviation radiophony concerned with the expertises on recordings, related to aviation incidents or accidents.

3 METHOD

The analysis of voice records of more than 70 in-flight mishaps was performed. It ought to have made clear a rather wide spectrum of questions, which in addition has changed to a certain extent within years. Their specification gives Table 1. As a rule, each expert opinion answered to several questions.

For acoustical analysis of sound recordings as much as possible samples of pilot's communications were treated, both from current and from critical parts of the flight. Every utterance was analysed on the Sona-Graph 6061 B Spectrum Analyser, using the method described by HECKER et al. (2). The mean pitch (Fo), voice diapason (D) and the expiration rate (ER) were determined for each communication. The output of speech per breath was expressed by means of the s.c. expiration rate formula (18):

$$ER = \frac{\text{number of syllables}}{\text{duration in ms}} \times 100$$

Besides all speech emboli and omissions, such as repeated words, words disturbed by stammering, slips of the tongue, leaving out words and the so-called hesitation phenomena were assessed, considering the for a long time known experience of tight relationship between mentioned phenomena and the anxiety (19).

4 RESULTS

Let us leave unnoticed the various demands on the psychoacoustic expertise enumerated in

Table 1, except for those, where the Accident Investigation Board raised a requisition for the appretiation of the impact of emotional activation upon the pilot's behaviour during the coping with an in-flight emergency. In one third of events the appretiation of pilot's emotional state and nearly in every seventh case also the type of crew's behaviour were requiered. With growing experience in the appraisal of their mutual affection all expert's accounts included the position over this affair.

The analysis of voice records, embracing 52 % of fatal accidents, 32 % of non-fatal incidents and 16 % of other critical in-flight events showed, that the pitch rises in average by 40 - 50 % above the baseline already in the first utterance, emitted by the pilot after the onset of sudden emergency. The gain in pitch can reach even more than 100 % in some individuals. With the rise of the pitch also the voice band width widens by 50 - 80 %. The extent of the pitch increase is not incidental to the cogency of situation, bringing about an intense emotional activation.

It turned out, that the information on an airman's situational awareness affection with emotional strain is mainly encoded in the temporal parameters of the speech, viz. in the expiration rate of speech. Expiration rate during emotional speech varies in a rather complicated manner. In subjects, coping with the situation in "cold blood" it grows, whereas the speech of subjects, whose behaviour is altered by psychoemotional inhibition, it regularly becomes retarded. The deviations in positive or negative direction reach about 15 - 30 %. However, it is rather its dynamic course with respects to the dynamism of the pitch, than absolute changes of expiration rate, which carry the relevant information. The antagonistic movement of the pitch in relation to the expiration rate, creating the so-called "open scissors pnehomenon" can be considered as a symptomatic manifestation of the situational awareness loss in a life-threatening situation.

Some typical illustrations of speech characteristics in subjects, who perceived the impassibilty of critical situation, are demonstrated on Fig. 1 and 2.

Fig. 1 depicts the voice characteristics in the reports of a young, inexperienced jet fighter pilot, who performed a series of unsuccessful attacs on a ground target. During the last, fifth attack one of the two airplane's engines stopped at a low altitude. Already in four communications, announcing the incorrect aiming on the target, the pitch and the expiration rate moved away. Still more conspicuous picture of escalating emotional stimulation and parallel intensification of helplessness occured during the 36 seconds long period of emergency. During the first 30 seconds the pilot presented signs of strong psychic and emotional inhibition. He resigned to passive execution of instructions, which he got from the flying controll officer. Gradual blockade of performance under the influence of asthenic emotion was reflected by a characteristic dissociation between increased pitch and decreased rate of speech. Only in the last 6 seconds, when the flying controll officer strictly ordered to leave the aircraft by ejection, pilot's voice reflected, along with acoustical signs of outlasting negative emotion, a reversal from pasive to active action, signalled by the expiration rate's rise.

Fig. 2 demonstrates an extreme rise of the pitch in the voice of a test pilot, whose turboprop lost the rudder. As he was fully aware of the incoclusiveness of the situation, the expiration

rate did not slow down excessively.

Fig. 3 brings an example of emotionally demanding emergency, which the pilot coped with appropriate awareness. After the evoking of one of the engine's pompage at safe altitude the pitch and expiration rate diverged only at the beginning of in-flight troubles, while in the course of deliberate solution of dangerous situation both speech characteristics moved in parallel. Immediately before the ejection pilot's voice again reflected a strong asthenic emotion, displayed by an opposite movement of temporal parameter.

5 DISCUSSION

The evaluation of behaviour of a pilot, posed with an in-flight emergency is a very demanding and responsible matter. Even a well trained flyer can appear face to face with a situation, which he/she hardly copes with. If such a situation ends in an accident, every information, which can lead to the more objective appretiation of what was really going on, is of immense value. Since late sixties the psychoacoustical analysis of air-to-ground communication contributes significantly to the precision of the crew's behaviour in mentioned events as well.

The establishment of the intensity of an airman's emotional activation gives only a very limited information of his/her actual performance under psychological stress. Sudden impairment of a planned flight mission with significant alteration of either airplane's run or pilot's ability to control the flight brings about the necessity of crucial transformation of perceptual, cognitive and motor skills. Intense emotion can enhance or block these activities.

The explanation of described changes of the speech signal characteristics must be derived from the relations between cognitive functions and emotional activation in stressful environment. The emergency significantly interferes with the maintaining of awareness of the tactical situation and not infrequently for once distracts the pilot from flying the aircraft. The rise of the pitch in stressful conditions should be considered as a natural projection of emotional activation through the mediation of speech organs. If the stimulation is faster than the start of adaptation, the pitch, due to the intense strain of the vocal cords can reach extreme values (see Fig. 2). Exaggerated emotional stimulation at the same time creates the tension of voluntary muscles of the trunk, which in jeopardy of one's life is perceived as a whole body "stiffness". Explicit stiffness of the respiratory and phonation muscles slows down the speech rate.

When cognitive processes, which are the core of operational awareness, prevail over the affective reaction of the pilot, the speech index parametres from now on move parallel to one another. Except for the terminal poles of emotional continuum the speech rate does not reflect merely the dynamism of psychological processes in an operator's mind. Already 13 years ago significant relations between the output of speech per breath (i.e. the expiration rate) and the semantic content of communication, accompanying ordinary air traffic operations, were proved (20). Highly formalized information messages, produced by pilots and ATC officers are produced in more slower rate, than interrogative utterances (Fig. 4). Regardless of the semantic content shorter utterances are produced more slowly, than the longer ones. For all that the speech fluency remains balanced, obviously thanks to the unconscious equalization of formal length differences by means of the speech expiration economy. The loss of operational

awareness and the predominance of emotional inducements over the rational solving of complicated in-flight tasks unbinds the interrelations between structural and formal components of speech.

6 CONCLUSIONS

The psychoacoustic and psycholinguistic analysis of an air-to-ground communication, performed in the course of coping with in-flight emergencies, can be of great value in an appraisal of pilot's situational awareness. Understandably, to embrace such an opportunity can only those, who are fully acquianted with the theoretical basis and practice of psycholinguistics, and the rules of dyadic communication as well. The new generation of speech analysers substantially simplifies the technical aspect of the speech signal processing. Irrespective of the impressiveness of the results of acoustical analysis their interpretation cannot be made without serious appreciation of other information, concerning the pilot's behaviour. Besides vocal indicators of psychological stress also other important speech signs of emotional activation, as the hesitation patterns, slips of the tongue, syntactic alterations etc., must be taken into account, just as the results of bio/histochemical postmortem analyses.

Even in the tight scope of the air-to-groud communication formal rules, the ground partner of a pilot, who copes with an emergency, can render a significant help to the aviator by fulfilling his/her demand by means of an activising communication.

7 REFERENCES

1. Essen,O.von: Psychophonetik. Arch. Ohr. Nas. Kehl. Heilkunde, 1954, 165, pp. 595 - 600.
2. Hecker,M.H.L. et al.: Manifestation of task-induced stress in the acoustic speech signal. J. Acoust. Soc. Amer., 1968, 44, No. 4, pp. 993 - 1001.
3. Williams,C.E., Stevens,K.N.: On determining the emotional state of pilots during flight: an exploratory study. Aerospace Med., 1969, 40, No. 12, pp. 1369-1372.
4. Williams,C.E., Stevens,K.N.: Emotions and speech: some acoustical correlates. J. Acoust. Soc. Amer. 1972, 52, No. 4/2, pp. 1238-1250.
5. Popov, V.A. et al.: Assessment of emotional strain of a cosmonaut during cosmic walk by means of physical parameters of speech. Proc. 15th Congress of Aviat. and Space Medicine, Prague, October 1966.
6. Popov,V.A. et al.: Analysis of intonational speech characteristics as an indicator of emotional state during the space flight. (In Russian). Zhur. vys. nerv. Deyat., 1966, 16, No. 6, pp. 974-983.
7. Popov,V.A. et al.: Speech frequency spectrum as an indicator of the level and character of emotional state in a man. (In Russian). Zhur. vys. nerv. Deyat., 1971, 21, No. 1, pp. 104 - 109.
8. Simonov,P., Frolov,M.V.: Utilization of human voice for estimation of man's emotional stress and state of attention. Aerospace Med., 1973, 44, No. 3, pp. 256-258.
9. Niwa,K.: Changes in voice characteristics in urgent situations. I. Rpt. Aeromed. Lab JASDF, 1970, 11, No. 1, pp. 51-58.
10. Niwa,K.: Change of voice characteristics in urgent situations. II. Rpt. Aero-

med. Lab JASDF, 1971, 12, No. 3, pp. 246-251.

11. Šulc,J.: Quantitative and qualitative characteristics of speech reaction of an operator in stressful conditions. Activ. nerv. sup. (Praha), 1973, 15, No. 1, pp. 50-51.

12. Šulc,J.: Problematics of radiophony from the point of view of aviation medicine. (In Czech). Voj. zdrav. Listy, 1974, 44, No. 2, pp. 41-44.

13. Šulc,J.: Voice as an indicator of man's condition in psychical and physical stress. (In Czech). PhD Thesis. Prague, Charles University 1974.

14. Šulc,J.: Expertise value of psychogenic changes in voice and speech. (In Czech). Voj. zdrav. Listy, 1975, 44 No. 6, pp. 226-229.

15. Šulc,J.: Projection of physical flight stressors in the voice and speech. (In Czech). Voj. zdrav. Listy, 1975, 44, No. 6, pp. 230-233.

16. Šulc,J.: Significance of nonverbal characteristics of speech signal to assess the psychophysiological state of a pilot. (In Russian) Kosm. Biol. Aviakosm. Med., 1976, 10, No. 6, pp. 54-58.

17. Šulc,J., Morávek,M.: Analysis of verbal behaviour of the first Czechoslovak cosmonaut during space flight. Cs. Psychol., 1979, 23, No. 1, pp. 42-49.

18. Goldman-Eisler,F.: Psycholinguistics. London, Academic Press 1968. 169 pp.

19. Kasl,S.N., Mahl,G.F.: The relationship of disturbances and hesitations in spontaneous speech to anxiety. J. Person. Soc. Psych., 1965, No. 1, pp. 425-433.

20. Šulc,J.: The dependence of speech rate on the semantic level of communication. Activ. nerv. sup. (Praha), 1982, 24, No. 3, pp. 171-172.

Table 1. Requirements for the psychoacoustical analysis of radiophony recordings

REQUIREMENT	SATISFIED	PART.SATISF.	NONSATISF.	TOTAL[%]
TYPE OF EMOTION	31.3	2.2		33.6
TYPE OF BEHAVIOUR	14.9			14.9
EXACT PHONETIC TRANSC.	17.2	5.2	5.2	27.6
TIME ANALYSIS	9.7	0.8	0.8	11.3
SPEAK./ TRANS.IDENTIFIC.	8.9	1.5		10.4
NOISE SPECTRUM	2.2			2.2
TOTAL	84.2	9.7	6.0	100.0

Fig. 1. "Open scissors phenomenon" in the reports of a disconcerted pilot, posed with an emergency (see text)

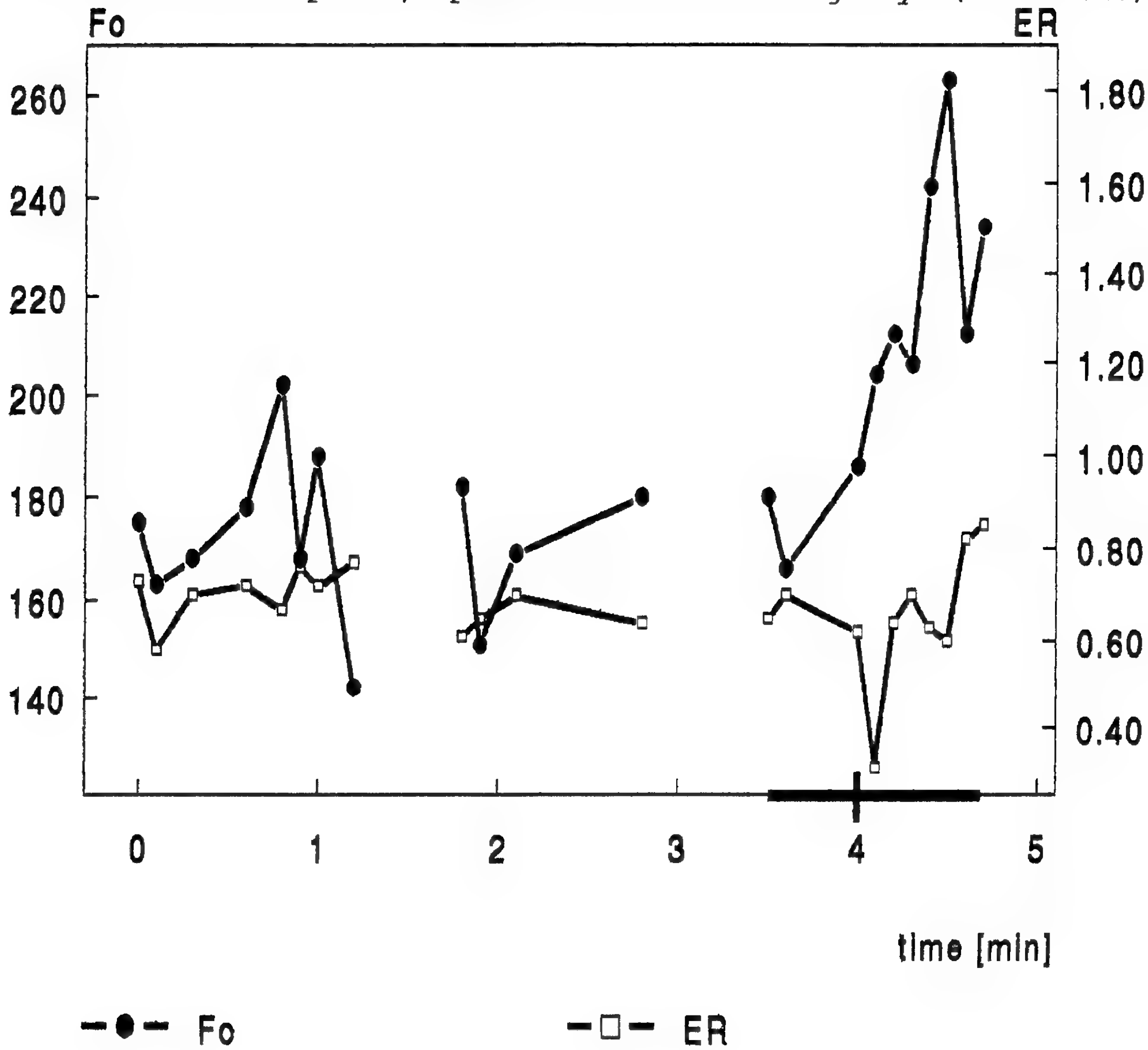


Fig. 2. Extreme rise of the pitch in the voice of a pilot, exposed to a hopeless situation (see text)

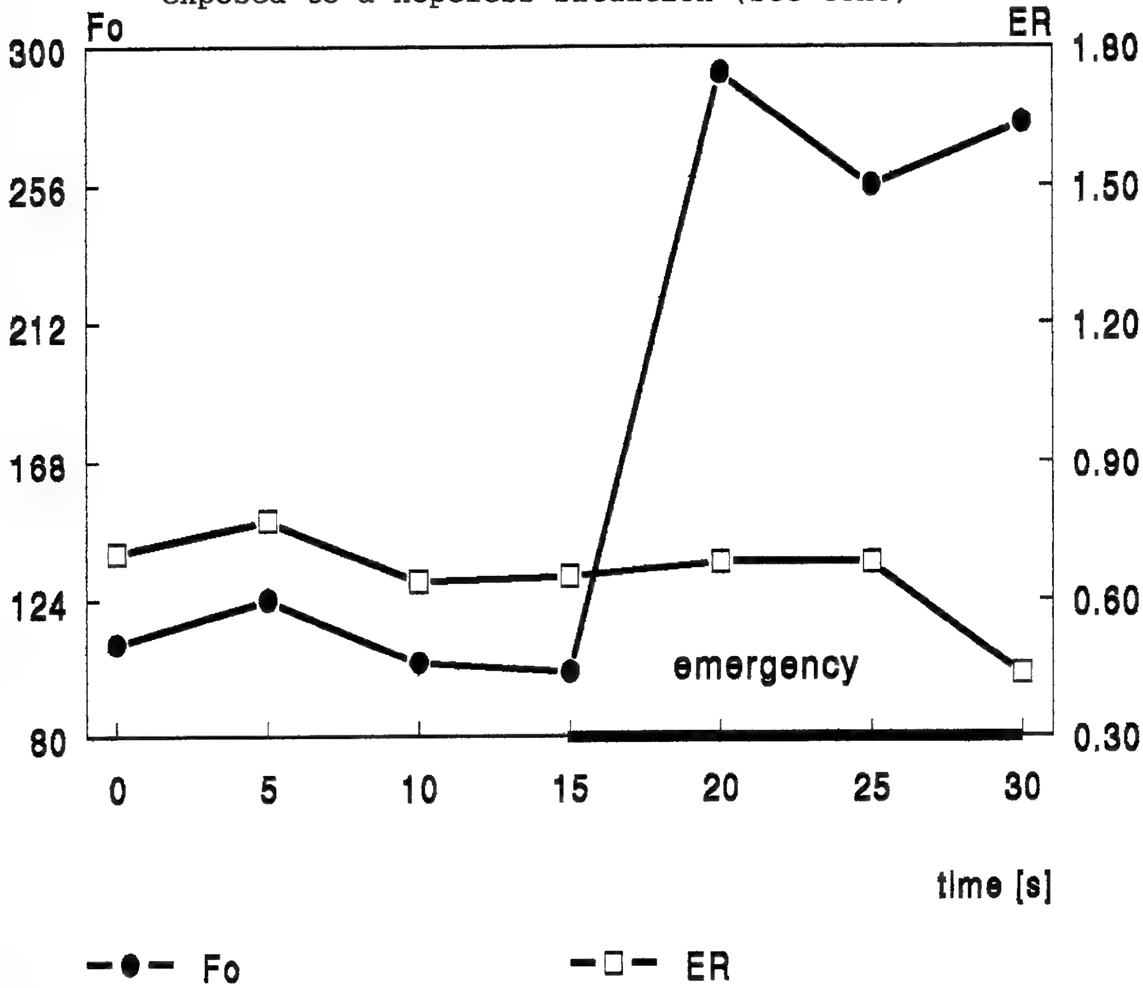


Fig. 3. Fo and ER in utterances of a pilot, managing an in-flight emergency with appropriate awareness (see text)

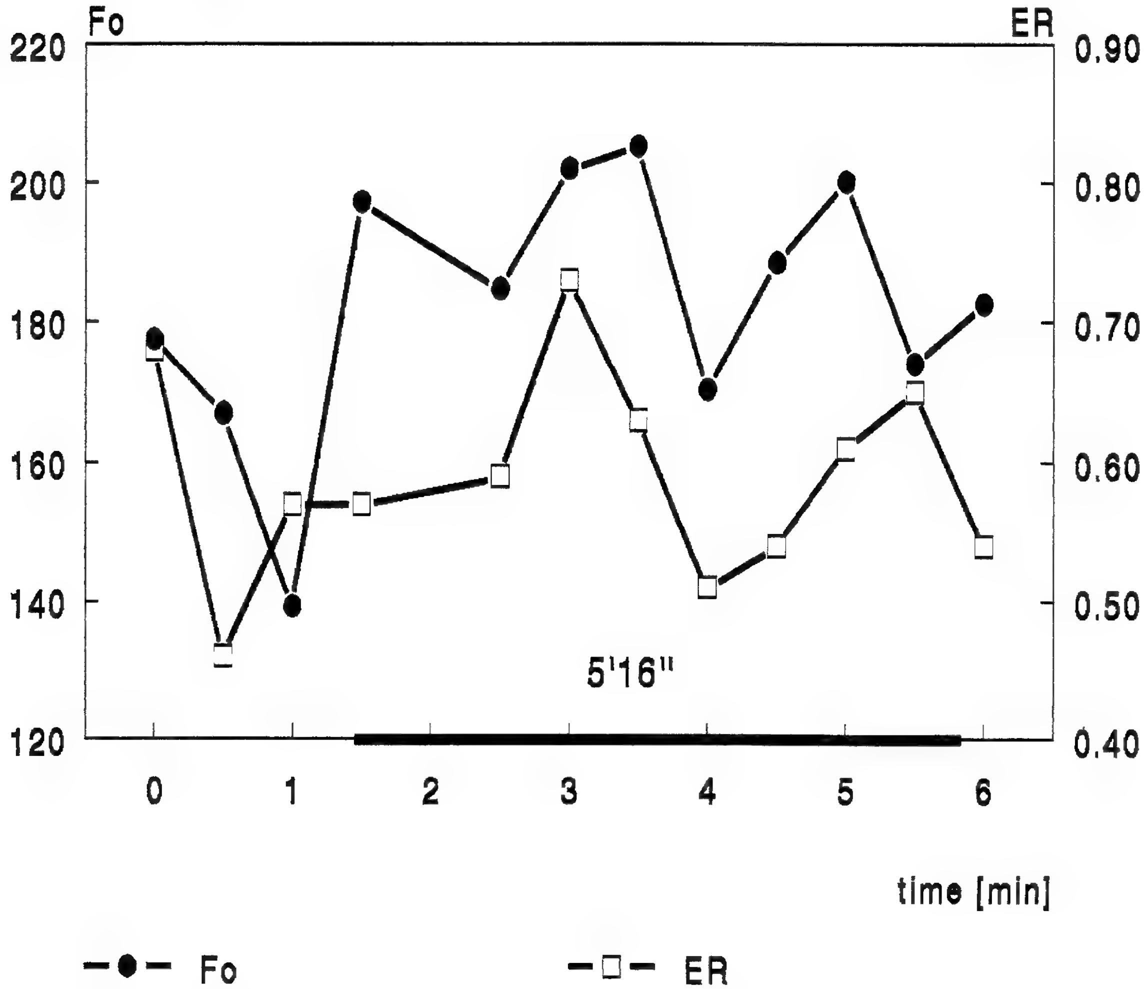
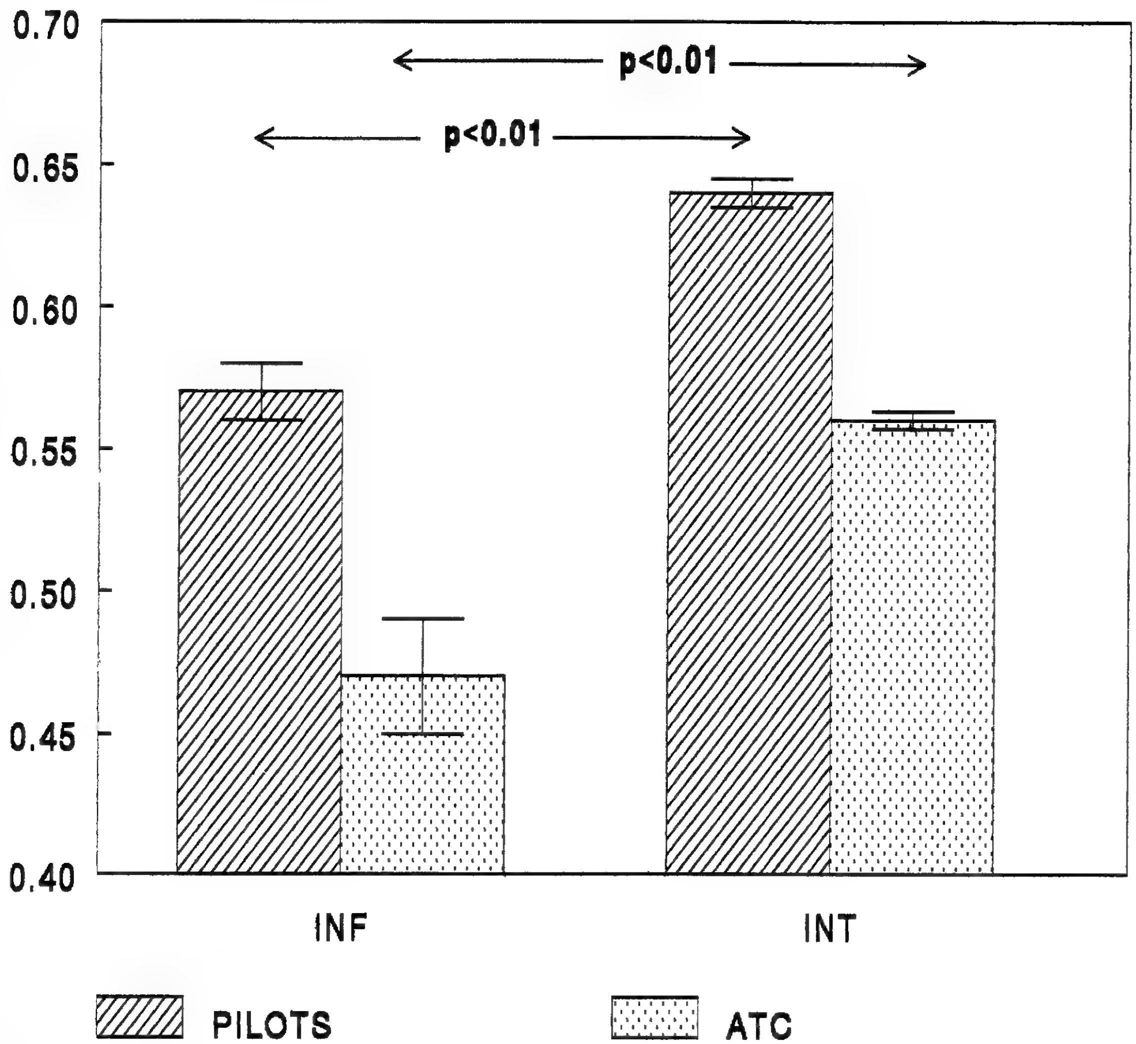


Fig. 4. Differences of ER in informatory and interrogative utterances



Viseur de casque et amélioration de la perceptionde la situation spatiale: Approche expérimentale de l'intérêt et des limitations.

A. LEGER, J.P. CURSOLLE, F. LEPPERT

SEXTANT Avionique, Military Avionics Division
Rue Toussaint Catros
BP 91
33166 Saint-Médard-en-Jalles Cedex
Le Haillan, FRANCE

SUMMARY: Enhancing Spatial situation awareness in HMDs: An approach of interest and limitations through experimental results.

Maintaining spatial situation awareness in modern fighter aircraft is generally considered as a challenging issue. In regards of recent progress made in HUDs formats, HMDs format requirements appear far from being clearly established. Encouraging results obtained by Osgood pointed out the potential interest of off-boresight symbology for low level flying and ground attack missions. Such symbology could be of considerable interest to enhance in HMDs the usually poor quality imagery (I^2, IR) used during night attack mission.

A binocular night attack HMD has been developed by SEXTANT on an exploratory development launched by the French DGA in 1991. Part of this development was devoted to definition and implementation of a set of symbology to be used in conjunction with imagery during flight tests on a Mirage 2000 test-bed Aircraft at the Flight Test Center(Brétigny sur Orge). To achieve this goal, an experiment was carried out using the equipment developed for the flight tests.

In a first experiment, short operational scenarios including low level navigation through mountains, runway attack and ground threats escape maneuvers were flown by experienced test and military pilots. Symbology and imagery were generated by a Silicon graphics "Onyx" graphic workstation. Symbolic and sensor imagery presentations were slaved to head movement and the pilot was fully in control of aircraft maneuvers. A virtual immersion technique was used to compare conformal and non-conformal formats (attitude and trajectory). Results showed that most pilots recognized the potential interest of conformal symbology, but also that the format used ("bird cage") was too disorienting to be really useful. Though non-conformal symbology was seen as a rupture in the logic of presentation between HUD and HMD, it was also felt that, provided some improvements were made, it remained the most readily acceptable format.

Lessons learned from the first experiment have conducted to an improved set of symbology, validated further on with semi-virtual immersion techniques. More general conclusions regarding methodological aspects of assessing the enhancement brought to imagery by off-boresight symbology and perceived limits of the formats used were also drawn

1. INTRODUCTION

Un potentiel d'amélioration de la perception de la situation a souvent été attribué a priori aux viseurs de casque pour les pilotes d'avions de combat. L'exploitation réelle de ce potentiel nécessite en premier lieu une maîtrise satisfaisante de tous les éléments ergonomiques de conception imposés par les contraintes d'environnement (masse, centre de gravité, encombrement).Une fois cette étape franchie, la question se pose alors de savoir jusqu'où peut aller cette amélioration, en particulier dans le domaine de la perception spatiale (de soi-même et de l'avion dans l'espace). La démonstration expérimentale des avantages éventuels des équipements, en simulateur ou en vol, est d'autant plus délicate que s'impose ici le principe de base de la médecine "primum non nocere". Les pilotes sont d'ailleurs particulièrement attentifs à la possibilité d'un risque de désorientation induit par une symbologie inadéquate.

L'avantage fondamental du viseur de casque (VDC) est qu'il permet de dépasser les limitations inhérentes au champ réduit de présentation des informations du collimateur tête haute (CTH). Le VDC offre au pilote la possibilité de consulter hors du champ du CTH des informations non conformes (états du système) ou des information conformes (à la précision près des capteurs). Cette caractéristique permet d'exploiter pleinement les capacités des systèmes d'armes modernes dans la totalité leur domaine(figure 1) (2).

Pour ce qui concerne le combat Air-Air, l'intérêt de la désignation d'une cible hors champ du CTH à fait l'objet en France de plusieurs études en simulateur de combat (1) dont les résultats ont conduit à l'adoption du VDC dans le cadre du programme Rafale. Les essais en vol réalisés dans le cadre du développement de l'équipement

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

pour ce programme ont permis de souligner deux avantages très appréciés des pilotes. Ces avantages sont essentiellement liés à la possibilité de visualiser d'une manière conforme dans le VDC la position d'une cible acquise par les capteurs du système. Ceci se traduit quantitativement par une amélioration de la distance d'acquisition visuelle qui, selon les conditions météorologiques, peut s'accroître de près de 50%. La possibilité de retrouver facilement une cible une fois qu'elle ait été acquise permet également au pilote d'accroître son niveau de confiance en combat

En revanche, la nécessité d'avoir en combat aérien une symbologie d'orientation spatiale (attitude , pente, trajectoire) demeure un sujet de discussion parmi les pilotes. Si le besoin d'une aide à la récupération de situation inusuelle est relativement admis, le type d'information à présenter n'apparaît pas très clairement. En dépit des résultats positifs obtenus par Osgood et Coll. (5, 10) il semble encore exister dans ce domaine une incertitude appelant des travaux de clarification et de définition de concept.

Les viseurs de casques de troisième génération, capables de présenter aussi bien de l'imagerie capteur que de la symbologie, offrent de nouvelles possibilités en terme de perception de la situation dans le contexte des opérations nocturnes en basse altitude ou en condition de mauvaise visibilité.

A l'heure actuelle, un certain nombre d'appareils modernes effectuant de nuit des opérations Air-Sol sont dotés de capteurs d'imagerie IR (FLIR) présentant l'image du terrain ou de l'objectif, soit sur un écran tête basse, soit dans le CTH. Des jumelles de vision nocturnes sont parfois associées à ces visualisations. En dépit de d'aspects indéniablement positifs, ainsi que le souligne Evans (4), ce type de présentation d'information est loin d'être optimal, pose des problèmes techniques et résulte généralement en une amélioration de la perception de la situation toute relative.

L'intérêt de capteurs IR dont l'orientation est asservie au mouvement de la tête couplés à un viseur de casque a été montré récemment, en particulier par Lydick et Hale et Coll (9, 6). Bien que les résultats obtenus apparaissent encourageants sur le plan de la perception de la situation spatiale, les possibilités de présentation d'imagerie demeurent encore relativement limitées dans un avion de combat. La suppléance visuelle ainsi réalisée est loin d'atteindre les caractéristiques de la scène visuelle perçue par le pilote lors d'une mission par ciel clair. Les limitations des capteurs d'imagerie, les contraintes biomécaniques entraînées par le port des équipements de tête dans un avion de combat et les solutions optiques compatibles des exigences opérationnelles contraignent sévèrement les possibilités dans ce domaine.

Dans la mesure où il semble difficile dans un futur proche de pouvoir disposer d'équipements avionnables offrant un très grand champ de vision (de l'ordre de 120°) et présentant une imagerie couleur haute résolution, il semble improbable qu'une perception totalement satisfaisante de la situation spatiale, reposant uniquement sur l'imagerie, puisse être obtenue. Le problème qui se pose est alors de savoir si la superposition d'information symboliques sur l'imagerie, ce qui constitue une fonction de base des visuels de 3ème génération, est susceptible d'améliorer cette perception avec les équipements actuellement disponibles.

2. CONTEXTE DE L'ETUDE

Un développement exploratoire concernant la réalisation d'un viseur de casque de troisième génération pour avion d'armes à été lancé dès 1990 par la Délégation Générale pour l'Armement (DCAé/ STTE/ DCT.4). La description détaillée de l'équipement réalisé (dit "Grand Champ Avion") et les différents compromis qui ont du être réalisés ont été présentés par ailleurs (8, 2).

Dans le cadre de ce développement exploratoire, la mise au point d'une symbologie comportant des éléments d'aide à la perception de la situation (APS) tactique et spatiale a fait l'objet de l'étude expérimentale présentée dans ce document. Cette symbologie, utilisée dans un premier temps avec une imagerie de qualité moyenne obtenue par la caméra ILCCD intégrée au casque, est destinée à la réalisation d'essais en vol sur le Mirage 2000 "BOB" (Banc Optronique Biplace) du Centre d'Essais en Vol (CEV) de Brétigny sur Orge.

Au delà de la mise au point d'une symbologie "d'orientation spatiale" destinée aux essais en vol, le but de l'expérimentation a consisté dans un premier temps en une évaluation comparative de deux concepts de symbologie pour le visuel de casque "grand champ avion", symbologie "synthétique" et symbologie "conforme". Une seconde phase expérimentale à été ensuite consacrée au développement du jeu de symbologie de base, en tenant compte des résultats obtenus précédemment.

3. METHODES

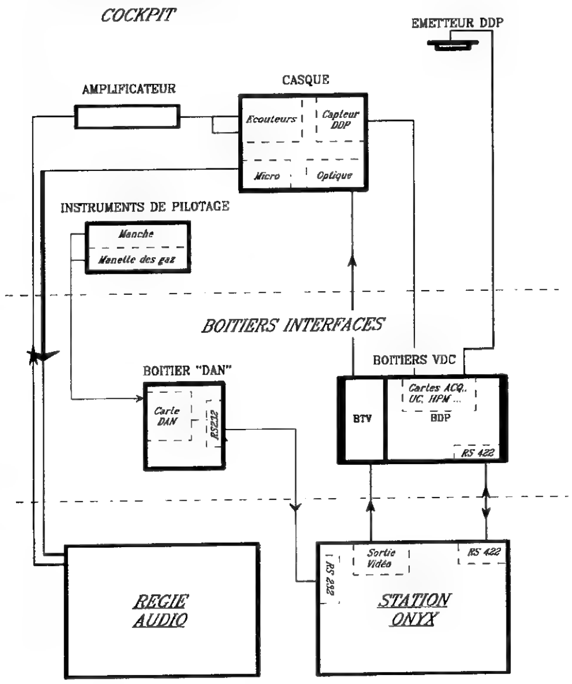
3.1. sujets

Six pilotes ont participé aux expérimentations (3 pilotes d'essais CEV et 3 pilotes du Centre d'Expérimentations Aériennes Militaires (CEAM)). Tous ces pilotes avaient une connaissance préalable des systèmes de vision nocturne.

3.2. Environnement matériel et logiciel:

Il est commun aux deux phases expérimentales. Le synoptique de l'installation est présenté dans la figure 2. Le dispositif s'articule autour d'une station de travail Silicon Graphics "ONYX". Les images générées sont asservies au mouvement de la tête, le pilote étant par ailleurs en contrôle des évolutions du modèle avion de simulation. Elles sont présentées en monochrome vert, dans un champ de 40X30. La qualité des images a été jugée meilleure que ce qui est actuellement généré par les capteurs réels (caméra thermique ou JVN), mais l'ensemble de la simulation a été globalement jugé

comme suffisamment représentative et acceptable pour les objectifs poursuivis.



Les pilotes étaient installées dans une cabine de Mirage 2000 sur un siège MK 10 . Les manettes de contrôle (manche et gaz) étaient simplifiées et non représentatives des dispositifs réels. La cabine ne comportant aucun instrument, les pilotes ont été placé dans une situation d'immersion virtuelle complète. Une visière totalement opaque était alors utilisée. Une symbologie classique CTH M 2000 apparaissait dans l'axe de la cabine et était remplacée par la symbologie périphérique lorsque la tête du pilote s'éloignait de l'axe du fuselage. Au cours de la seconde phase expérimentale, cette situation d'immersion a été complétée par la réalisation d'une condition de semi-immersion. Dans ce cas, le pilote peut voir les structures de la cabine au travers de la visière, une toile noire opaque isolant la cabine dans le local d'expérimentation.

3.3. Scénarios opérationnels:

Ils ont été définis par les experts opérationnels pilotes de SEXTANT qui ont également participé à la mise au point des expérimentation.

Trois scénarios de base ont été retenus:

A : Pénétration BA - attaque d'objectif - dégagement
B : A+ passage IMC (sans visibilité) au point clé haut.
C : Navigation en zone hostile avec menaces.

3.4. Déroulement de l'expérimentation

3.4.1. Consignes préliminaires:

Avant de commencer la séance de travail, le but de l'expérimentation était exposé aux pilotes expérimentateurs. Le déroulement de l'expérimentation était commenté et une présentation statique des différents éléments de symbologie était effectué.

Pour l'exécution des scénarios de mission, il était précisé que les consignes d'altitude et de vitesse devaient être respectées du mieux possible, mais que l'achèvement complet de la mission est prioritaire, quelles que soient les conditions rencontrées (en particulier passages en IMC).

3.4.2 Entraînement

L'entraînement comportait 3 phases

- Evolutions libres (comportant au moins un atterrissage). Il s'agissait d'une prise en main des contrôles et d'une accoutumance aux différentes caractéristiques de la simulation..
- Présentation en dynamique de la symbologie synthétique, évolutions libres (durée selon demande pilotes).
- Présentation en dynamique de la symbologie conforme, évolutions libres (durée selon demande pilotes).

A la fin de l'entraînement, on s'assurait que le pilote avait une connaissance suffisante de l'environnement de simulation et qu'il maîtrisait les caractéristiques de pilotabilité, ainsi que la signification des différents symboles.

3.4.3 Plan d'expérience

première phase expérimentale:

Tous les scénarios étaient réalisés par chaque pilote, chaque pilote étant son propre témoin. Trois conditions de symbologie viseur de casque étaient utilisées pour chaque scénario, synthétique (SY), conforme (CF), pas de symbologie (N).

En tout, chaque pilote effectuait 9 essais résultants de la combinaison des 3 scénarios (A, B, C) avec les trois symbologies. L'ordre de présentation des combinaisons était réalisé selon un plan type "carré latin", afin d'éviter les effets d'ordre.

En fait, si ce plan d'essai a bien été suivi dans l'ensemble, certains pilotes n'ont pu effectuer la totalité des essais, en particulier pour ce qui concerne le scénario C. Le déroulement des session a été réaménagé en conséquence.

Deuxième phase expérimentale:

La deuxième phase expérimentale a été réalisée selon un protocole allégé, le scénario C n'étant pas utilisé.

Cette phase a uniquement porté sur l'évaluation de différentes alternatives de la symbologie conforme, modifiée en fonction des besoins exprimés par les pilotes dans la première phase.

les sources de variations portaient sur le graphisme de l'indicateur montée/ descente/attitude et sur la présentation des informations d'altitude et de vitesse (cadrans ou alphanumériques).

3.5. Données recueillies

Variables quantitatives:

Profil de trajectoire en altitude, profil de vitesse Taux maximum de virage(Θ) aux points clés et lors des manoeuvre d'évitement.

Données qualitatives:

Un questionnaire d'évaluation était présenté au pilote à l'issue de chaque passe d'évaluation. Le contenu du questionnaire a été déterminé avec la participation des pilotes experts de SEXTANT. Une échelle d'évaluation comportant 4 niveaux, inspirée de travaux menés au Centre d'Essais en Vol, était utilisée.

Afin de compléter ce questionnaire sur le plan de la perception de la situation un observateur, expert pilote de SEXTANT, réalisait systématiquement une évaluation de l'état du pilote expérimentateur, en particulier dans les phases comportant des évolutions rapides.

4. SYNTHESE DES RESULTATS ET DISCUSSION

Les résultats obtenus portent sur différents points: validité de la méthode, symbologie conforme/non conforme, symbologie de base.

4.1. Validité de la méthode:

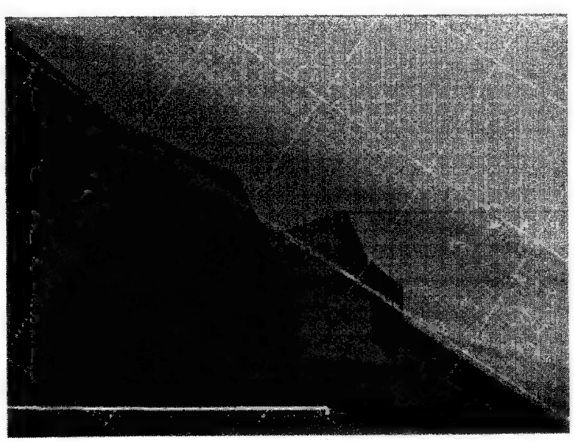
En dépit de la grande "rusticité" de l'environnement de simulation, les pilotes expérimentateurs ont dans l'ensemble jugé que la technique utilisée permettait d'atteindre les buts fixés (évaluation de concept). Les points faibles de la simulation (modèle avion, image capteur, traînage de la symbologie et qualité du graphisme) ont été reconnus, mais, dans le contexte fixé, sont apparus suffisamment acceptables pour permettre de travailler efficacement. La grande flexibilité apportée par l'utilisation d'une station de travail couplée au viseur de casque a par ailleurs été appréciée, dans la mesure où elle permettait d'accéder presque "en temps réel" à la présentation de modifications demandées par le pilote.

La situation d'immersion totale, même si elle a été globalement bien acceptée, pose plus de problèmes. En effet, il apparaît les informations proprioceptives au niveau du cou ne permettent que très imparfaitement la détermination de la position de la tête par rapport au repère cabine. Dans ce contexte d'immersion totale, l'information visuelle liée à la tête semble en effet excessivement dominante par rapport aux informations proprioceptives. Cette caractéristique, accentuée ou même intimement liée à la dimension restreinte du champ de vision, introduit des interférences néfastes sur la tâche de pilotage, en particulier lorsqu'il devient nécessaire de revenir rapidement dans les informations stabilisées du CTH ou pour l'évaluation de la distance angulaire entre le nez de l'avion et la direction de la visée en périphérie. En raison de l'opacité de la visière, la perte de la vision périphérique et l'inefficacité fonctionnelle des déplacement de l'oeil au delà de ± 20 constituent ici un élément déterminant. Cette situation n'est bien sur pas représentative de la situation d'emploi prévue en vol, qui est beaucoup plus proche de la situation de semi-immersion réalisée lors de la deuxième phase expérimentale.

4.2. Comparaison symbologie conforme et non conforme:

Alors que la symbologie non-conforme avait pour base les réflexions menées dans le cadre de la mise au point d'une symbologie "Air-Air", la solution de symbologie conforme s'est inspirée des travaux menés par la DRA et présentés récemment par Ineson (7) ainsi que par Doyle (3). Le concept "Bird cage" a été utilisé avec la représentation d'un élément de structure appartenant à l'avion (aile virtuelle). L'objectif recherché étant essentiellement une aide à la perception de l'attitude plus qu'au pilotage proprement dit, aucune indication d'altitude, de vitesse ou d'énergie n'était présentée dans ce dernier cas (fig. 3).

En dépit de son caractère novateur, très éloigné des concepts classiques, la symbologie du type "bird cage" n'a pas été rejetée d'emblée par les pilotes. Un potentiel certain a été reconnu à ce type de concept, même avec la relativement faible dimension du champ du viseur. Il est cependant apparu assez rapidement que la précision requise pour évoluer en basse altitude ne pouvait correctement être atteinte avec la simple notion d'attitude fournie par la "bird cage" et l'aile virtuelle. Des ambiguïtés dans l'indication de montée/descente ont également été relevées lorsque l'inclinaison était forte. Mais surtout, dès que les évolutions devenaient rapides, un aspect très désorientant de ce type de symbologie est apparu, lié au traînage et au défilement des lignes, amenant à la constatation que la situation devenait "pire que rien" alors qu'elle était plutôt jugée "mieux que rien" avec la symbologie synthétique.



Cette dernière attirait également des critiques relativement sévères, aussi bien sur le fond que sur la forme. En particulier, la rupture de logique de présentation des repères d'attitude et de pente entre le CTH et la symbologie périphérique a posé un problème à certains pilotes. A elle seule, cette critique n'apparaissait pas vraiment rédhibitoire, mais les défauts de forme existant dans le jeu initial de symbologie synthétique rendaient clairement cette figuration inacceptable pour les évolutions en très basse altitude.

D'un autre coté, il est apparu que le champ de vision limité et la qualité relativement pauvre de l'image du paysage, comportant en particulier très peu de détails de texture, ne permettait pas non plus d'exploiter les potentialités offertes par le viseur de casque, à l'exception de la fonction désignation.

Les principales remarques faites sur la symbologie synthétique étaient les suivante:

- taille globalement trop réduite de la symbologie
- Manque de précision dans la détermination de la pente
- manque d'information en altitude et vitesse
- manque d'information sur la cadence des évolutions dans le plan horizontal.

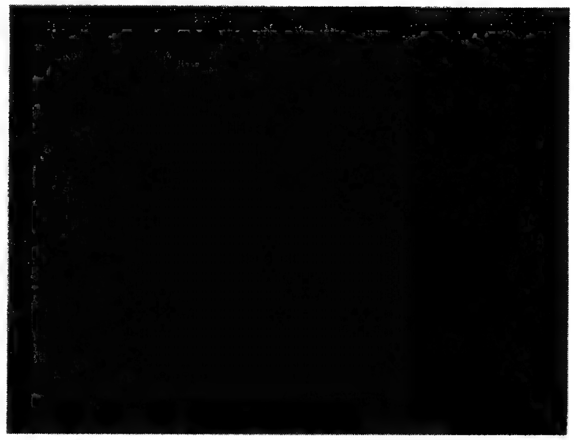
A l'issue de la première phase expérimental il apparaissait donc que l'existence d'une symbologie périphérique superposée à l'imagerie était bien nécessaire pour évoluer en très basse altitude et exploiter les possibilités du viseur de casque. La symbologie synthétique, bien que présentant quelques défauts inacceptables sur la forme, est apparue comme la plus susceptible de pouvoir conduire à court terme à la réalisation un jeu de symbologie utilisable pour des essais en vol.

4.3. Symbologie de base pour les essais en vol

Le développement du jeu de symbologie pour la réalisation des essais en vol a donc exclusivement consisté en l'amélioration du jeu initial de symbologie synthétique.

Il est assez intéressant de constater que les évolutions amenées à partir des critiques des pilotes expérimentateurs a conduit a réaliser une symbologie représentative de la notion de "T-Basic" (fig. 4). On peut ici se demander si ce résultat est lié à la culture des pilotes (instruits et entraînés sur cette notion de base), ou si en périphérie comme en axial, le concept du "T-Basic" est tellement robuste pour le contrôle du vol qu'il s'applique à toutes les situations. Parmi les alternatives de présentation d'information, l'utilisation de cadrans, qui donnent une bonne notion du sens de variation, a été largement préférée aux indications numériques.

Pour ce qui concerne l'indicateur montée/descente, on a constaté qu'une forme de présentation trop proche de l'échelle de tangage du CTH (fig. 5) pouvait entraîner une confusion.



Dans ce cas et particulièrement sous forte contrainte temporelle, le premier réflexe pour se mettre en montée consiste à lever la tête, ce qui a pour effet de mettre l'échelle et la maquette avion dans le ciel. L'action appropriée su la profondeur n'intervient que secondairement, après que l'absence d'effet sur la trajectoire de l'avion ait été reconnue. Ce phénomène apparaît beaucoup moins prononcé avec une échelle de montée/descente introduisant une rupture logique franche avec les informations du CTH.

Depuis la phase d'expérimentation initiale, la symbologie synthétique a pu être présenté à une large population de pilotes, en particulier dans le cadre d'expérimentations menées en coopération avec l'ARMSTRONG LABORATORY (WRIGHT-PATTERSON AFB) et le CERMA sur le couplage d'informations sonores localisées (Son 3D) et du viseur de casque. Les retours obtenus sont dans l'ensemble positifs et indiquent que l'adaptation à la symbologie périphérique s'effectue assez rapidement.

Les éléments de validation effectués dans la seconde phase expérimentale et des études suivantes montrent cependant assez clairement qu'un très bon contrôle des évolutions de l'appareil peut être obtenu lorsque le pilote utilise une symbologie périphérique convenable. Cette symbologie permet ainsi d'utiliser au mieux les fonctions de désignation et de présentation d'imagerie fournies par le viseur de casque. Un problème persiste cependant, plus au niveau de l'emploi que de la qualité des informations de pilotage. Il devient en effet extrêmement

facile et confortable de piloter l'appareil en regardant sur le côté, amenant parfois à oublier que les éléments de paysage qui sont vus ne sont pas ceux qui se trouvent sur la trajectoire de l'appareil. La stratégie d'utilisation du viseur de casque par le pilote doit donc être élaborée en fonction de cette notion. A terme, on pourrait coupler une fonction d'évitement de terrain pour apporter une réponse technique à ce problème. Ceci nous amène donc à considérer le perspectives ultérieures qui peuvent être envisagées pour l'utilisation d'un viseur de casque pour des missions nocturnes en basse altitude, utilisant une imagerie capteur et une symbologie périphérique.

4.3. Perspectives ultérieures

La symbologie développée pour les essais en vol a pour ambition de permettre au pilote d'exploiter au mieux une imagerie de qualité moyenne, comme celle résultant de la source IL intégrée au casque. L'approche retenue avec l'utilisation d'une symbologie synthétique non conforme est essentiellement fondée sur la notion "d'évitement de problème". Des essais complémentaires dans un contexte de récupération d'attitude inusuelle devraient permettre de tester la validité et la robustesse de cette symbologie en cas de problème avéré. Dans ce type de situation, certains éléments recueillis pendant la première phase expérimentale semblent indiquer qu'une symbologie conforme peut se révéler intéressante. Il semble donc opportun, parallèlement au essais de la symbologie synthétique, de poursuivre des études sur les bénéfices qui pourraient être retirés d'une symbologie conforme. Le premier point à examiner dans ce domaine est sans doute l'amélioration du graphisme et la connaissance des conditions limites à respecter pour éviter d'induire une désorientation là où une meilleure conscience de la situation spatiale est recherchée.

5. CONCLUSIONS

Les résultats obtenus lors de cette étude indiquent que, dans les conditions de simulation utilisées, l'enrichissement par la symbologie de l'imagerie présentée dans le champ du viseur de casque permettait un bon contrôle de la situation spatiale en vol basse altitude. Ce résultat doit bien sur être confirmé par les essais en vol avant de pouvoir être considéré comme acquis en environnement opérationnel. Il est cependant à craindre que le recours à des informations symboliques soit nécessaire pour obtenir une bonne perception de la situation spatiale tant que des progrès substantiels n'auront pas été réalisés en matière de résolution et de qualité d'image ainsi qu'en champ de vision.

Dans l'état actuel des connaissances, il est apparu que le concept d'APS fondé sur une représentation synthétique de la situation serait plus facilement utilisable pour les essais en vol qu'une représentation conforme du type "bird cage". Cette dernière représentation porte cependant un potentiel intéressant qui mérite d'être étudié plus en profondeur.

L'un des acquis intéressants de cette expérimentation est qu'un concept de symbologie considéré comme une solution "raisonnable" pour des essais en vol a pu être élaboré avec des coûts relativement bas dans un environnement de simulation très simplifié. Cela a pu être rendu possible grâce à la flexibilité des techniques de simulations utilisées, qui se prêtent bien à une expérimentation menée dans les règles des études de facteurs humains. L'utilisation d'un visuel de casque dont l'ergonomie est compatible avec les contraintes rencontrées en vol (en terme de masse, de centrage et d'encombrement) constitue sans doute un point important dans l'adhésion des pilotes à la méthode d'essai en "tâche partielle". Il doit être cependant clair que l'intégration dans un système complexe de ce type d'équipement nécessiter le recours à des techniques de simulation plus élaborées.

Enfin, les problèmes rencontrés dans l'utilisation d'une technique d'immersion totale dans l'environnement virtuel montrent bien les limitations de ce type de concept dans les applications embarquées et même vis à vis de la simple simulation de vol.

Remerciements: Cette étude à été financée par la DGA (DCAé/STTE/DCT.4, Mr CRIQUI). Les auteurs tiennent également à remercier J.C. FRIOT, C.GARDELLE, P. MOUNIN et J.M. MULTON pour leur contribution à cette étude.

Références bibliographiques

1- Bourse C., Demartial B. (1986): Campagne de combat ACT avec casque de visée: Analyse des résultats. Rapport N° ODSA/55017/STR 2/.293/CD

2- Cursolle J.P., Leger A., Leppert F., (1994): "Un visuel de casque pour la mission de nuit en basse altitude", MSP panel symposium Low Level And Nap Of The Earth (NOE) Night Operations, AGARD-CP 563, pp 22-1 to 22-10

3- Doyle A.J.R. (1994); Flight instrumentation for high alpha manoeuvres. Proceedings 4PWR SM TG Workshop on Human Factors/ future combat aircraft, Ottobrunn, volume 2, 21 p.

4- Evans L. "Fixed wing night attack: EO integration and sensor fusion," AMP panel symposium Helmet Monted Displays and Night Vision Goggles, AGARD CP-517, pp 1-1 to 1-5, Dec 1991

5- Geiselman E.E., Osgood R.K., (1994): Utility of off-boresight helmet-mounted symbology during a high angle airborne target accquisition task.

6- Hale R.A., Chino J.J., Niemyer L.L., Jadik J.R., Lightner B.E. "Design considerations for a night, air to surface attack capability on a dual role fighter," joint FMP/GCP panel's symposium Combat Automation for Airborne Weapon Systems: Man/Machine Interface Trends And

Technologies, AGARD CP-520, pp 30-1 to 30-8, Apr 1993

7- Ineson J.,(1993): The DRA virtual cockpit reseach program. AGARD-CP 541, AMP symposium on"Virtual interfaces: Research and applications. LIsbon, Portugal, 8,1-12

8- Leger A., Roumes C., Gardelle C., Cursolle J.P., Kraus J.M. (1994): Binocular HMD For Fixed-Wing Aircraft: A TradeOff Approach. Displays: 15;2, 91-97.

9- Lydick L.N. "Head-steered sensor flight test results and implications," joint FMP/GCP panel's symposium Combat Automation for Airborne Weapon Systems: Man/Machine Interface Trends And Technologies, AGARD CP-520, pp 15-1 to 15-7, Apr 1993

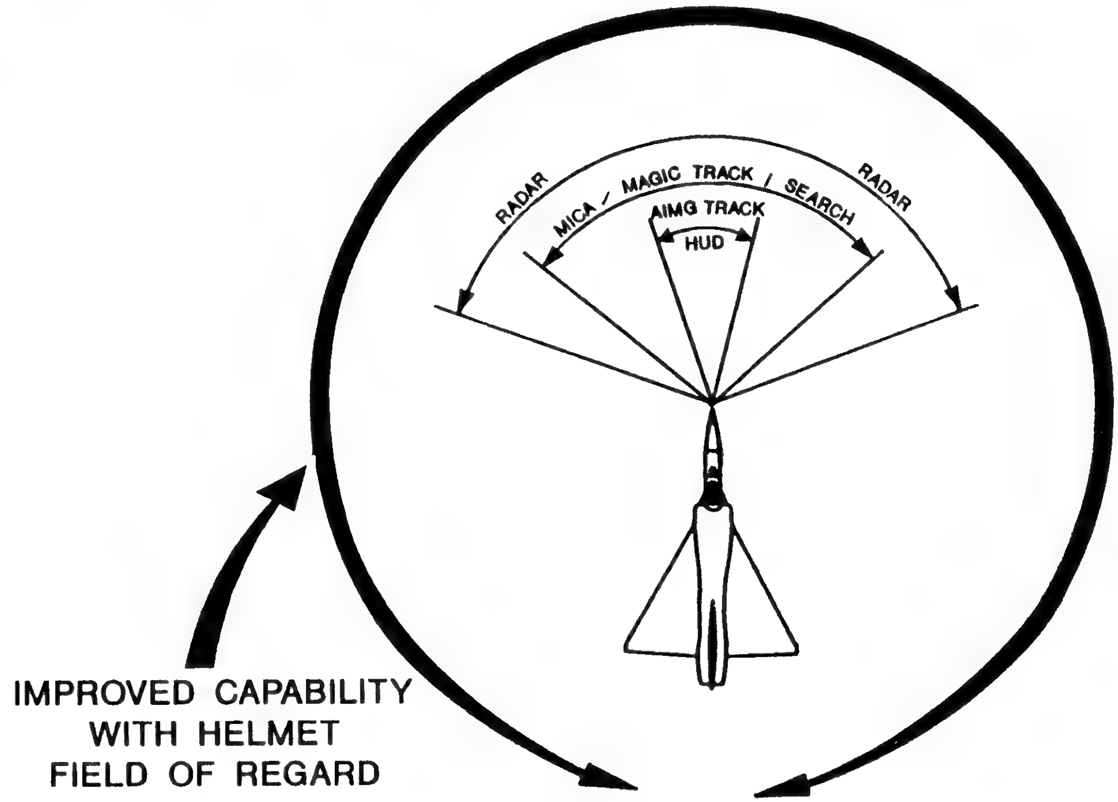
10- Osgood R.K., Geiselman E.E., Calhoun C.S., (1991): Attitude maintenance using an off-boresight helmet-mounted virtual display. AGARD-CP-517, AMP symposium on "Helmet Mounted Displays and Night Vision Goggles", Pensacola, FL, USA, 14, 1-7.

SEXTANT AVIONIQUE

MILITARY AVIONICS DIVISION

HUD LIMITATIONS

THE HUD FOV CANNOT MEET THE SYSTEM REQUIREMENTS



P193-619-7

Designing Novel Head-Up Displays to Promote Situational Awareness

T.D. Hardiman, H.J. Dudfield, S.J. Selcon and F.J. Smith
Human-Factors Group, Man-Machine Integration
Defence Research Agency, Farnborough, Hants GU14 6TD
United Kingdom

SUMMARY

This paper considers the design of attitude symbology for Head-Up Displays and describes two experimental studies conducted at DRA Farnborough. In these studies, novel and current HUD formats were compared in a range of tasks requiring attitude awareness. Both studies compared the novel formats in a fixed-based simulator using a 'recovery from unusual position' flight task and task-performance was measured. The trials differed in terms of the 'design driver' data taken. The first used workload ratings (NASA Task Load Index) and the second situational awareness (SA) ratings (Situational Awareness Rating Technique). Significant reaction time differences were found between the conditions in both studies. However, although these significant differences were supported by the SA ratings they were not reflected in the workload ratings. It is suggested that under certain trial conditions SA is a superior design driver to measure than workload, since workload does not include cognitive aspects of pilot performance such as prior knowledge and understanding.

1. INTRODUCTION

This paper discusses the merits of using situational awareness (SA) as a design driver in novel Head-Up Display (HUD) symbology design. It contains a description of two simulation evaluation trials. The primary purpose of the trials was to evaluate the provision of asymmetry on the HUD pitch ladder between the positive and negative pitch bars to reduce ambiguity. This concept is not new. The USAF standard HUD (MIL-STD-1787) has attempted to improve attitude referencing through the use of asymmetry between the bottom and top halves of the pitch ladder[1]. This is achieved through the provision of different shape coding for the negative (bendy bars) and positive (tapered) pitch bars. As an alternative to pitch bar shaping to provide asymmetry, pitch bar colouring has been suggested[2]. For the purposes of this paper, the trials provide an opportunity to examine the tools used in display evaluation. The data from these trials are reported in terms the utility of colour coding by Dudfield[3].

1.1 Task-Specific Performance Measures Versus Meta-Measurement of Design Drivers

Symbology evaluation trials typically consist of a subject performing an experimental flight task (e.g. recovery from unusual position), either in a simulator or in flight trials' resulting in measurement of the subject's behaviour. These measurements can be either task-specific measurements or the meta-measurement of design drivers.

Task-specific measures quantify specific, key aspects of human performance that ensure successful task completion. Examples of task-specific measures include speed, accuracy, root-mean square error (RMSE), eye movement and memory probes. These measures are described in greater detail by Hardiman, Dudfield, Newman, Doyle and Fearnside[4]. In general, these measures measure the cognitive, visual and psycho-motor performance of the subject when using symbology. These measures provide direct comparisons of symbology on parameters that are of direct relevance to specific flight tasks. However, a task-specific performance measure cannot by its very nature be a generic measure of symbology. Because these measures are task-specific, their relevance to the evaluation of symbology varies between experimental tasks. This has been demonstrated through inconsistencies and even contradictions in results between symbology evaluation studies. Further, reliance on task-specific measures prevents statistical comparisons between or across different studies being made as the task-specific performance measures are rarely the same. This results in the duplication of effort and fails to provide necessary guidance to symbology designers.

Meta-measurement offers an alternative yet complementary approach to task-specific performance measures. Meta-measurement is the measurement of design drivers. Design drivers are key human factors concepts that have been identified as generic rules or guidelines that should be followed in display design. For example, it is often cited that symbology should promote low workload or high situational awareness. In other words, design drivers predict that if certain rules or guidelines are followed in symbology design then performance benefits will result. These measures are 'higher level' or meta-measurements that can be applied across all evaluation tasks. This concept is not novel. Taylor and Selcon[5] describe SA as "meta-goal" which is not part of the mission goals but is necessary to allow mission goals to be attained. Meta-measurement of design drivers provides the possibility of a single measure being applied in evaluation trials to a range of displays, experimental tasks and situations. Further, design drivers allow a link between the design and evaluation stages in symbology development. Key design drivers can be identified as necessary requirements of a display in the design stage and measured in evaluation trials. For example, symbology designers can aim to design symbology that provides the pilot with a high level of situation awareness and actually measure the success of this in evaluation trials.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

A design driver should be recognisable by, and accessible between, different research organisations to encourage the comparison of formats that have been evaluated in different trials. The major design driver in aviation display design has traditionally been workload and more recently SA. The measurement of design drivers can either be taken in conjunction with task-specific performance measures or independently. When used in conjunction, the design drivers can be used to support task-specific performance measures. However, when the task-specific performance results are not supported or are even contradicted by the design driver measurement, then recommendations concerning the superiority of a symbology format are difficult to make. Alternatively, design drivers can be used as an alternative to task performance measures. However, this requires a design ruler that is a reliable predictor of performance.

2. EVALUATION EXPERIMENT 1

2.1 Workload as a Design Driver

Workload is often used as a 'design driver'; i.e. the attainment of a suitable level of pilot workload has driven the design of aircraft systems, in an attempt to reduce pilot error. The basic premise underlying the use of workload as a design driver in symbology evaluation is that the superior the symbology, the less workload it places upon the pilot. This concept itself is based on the theoretical basis that a certain amount of mental capacity of attention is available to the operator and that a display should optimise the use of that capacity[6]. Workload is affected by external variables (e.g. weather, external threats) and on-board variables (e.g. number of cockpit tasks, ease of understanding of displays). When workload is excessive, errors arise from the inability of the pilot to cope with high information rates and hence high attentional demand. When workload is too low, the pilot may become bored and may not attend to the mission tasks at hand, also leading to error.

Although workload has obvious relevance to the human factors of system design, it may be an insufficient design driver, since it concentrates on attentional demands without considering cognitive factors such as prior knowledge and understanding. As a design driver, predictions of the effect of workload on task-specific performance have had mixed success. One explanation of this is that workload is a multi-dimensional construct, that is effected by such a large number of variables, that it cannot be simply measured and quantified through techniques such as rating scales. An alternative explanation which must also be considered, however, is that workload is too narrow a concept to be useful in many cognitively complex aviation situations.

2.2 Method

Sixteen subjects (ten male and six female) completed the fixed-based simulation experiment. All were volunteers who had no previous experience of flying, and had normal or corrected to normal vision.

The experiment was a within-subjects design. Each subject completed 288 trials consisting of four conditions. The positive pitch bars were manipulated in each condition. The negative pitch bars were monochrome and tapered in all conditions. The conditions are depicted in Table 1.

Condition	*Positive Pitch Bar Shape*	*Positive Pitch Bar Colour*
No colour or shape asymmetry	Tapered	Monochrome
Colour asymmetry	Tapered	Blue
Shape asymmetry	Bendy	Monochrome
Colour & shape asymmetry	Bendy	Blue

Table 1: Conditions of Experiment 1.

The subjects were required to recover from unusual positions (UPs) as quickly and accurately (i.e. avoiding ground-collision) as possible. The 24 unusual positions were a combination of four roll angles (30°, 60°, 120° and 150°) and six pitch angles (70°, 50°, 30°, -30°, -50° and -70°). The order of the conditions was counterbalanced across subjects using a Latin Square, and the order of presentation of UPs was randomised.

The task-specific measures were initial reaction time (IRT) in ms (the time taken to make the initial stick movement), total recovery time (TRT) in ms (the time taken to recover to straight and level flight), the number of crashes (ground collisions) and the number of times each subject failed to recover within 20 seconds. Each subject also completed the NASA Task Load Index (TLX) workload questionnaire.

In the training stage, subjects were instructed on the use of HUDs and on the correct procedure for recovery from UPs. The subjects were told to roll and pull to the nearest horizon. They were allowed to practise the task until they felt comfortable with the recovery procedure and the format of the alternative HUDs. In the experimental stage, UPs were presented on a HUD display on a white background. Subjects completed all four conditions, providing TLX scores after each condition.

2.3 Results

2.3.1 Task-Specific Performance Data

The IRT, TRT, crash and failure to recover data were analysed using analysis of variance (ANOVA) with the

following fixed factors: condition, pitch angle, roll angle, order and repetition. All significant effects were tested post-hoc using the Newman-Keuls procedure. Significant differences were found between the HUD formats for IRT only. The task performance measures TRT, the number of crashes and failures to recover did not demonstrate differences between the conditions.

Subjects' IRTs differed significantly between the HUD formats ($F_{3,36}$=4.012, p<0.05). Subjects' IRTs were faster (p<0.05) for the colour and shape asymmetry condition (mean=949 ms) than for the no asymmetry condition (mean=1046 ms) and shape asymmetry condition (mean=1060 ms). Further, subjects' IRTs were lower in the colour and shape asymmetry condition than in the colour asymmetry condition alone (mean=990 ms). However, this result was not significant. The mean scores for each HUD condition are summarised in Figure 1.

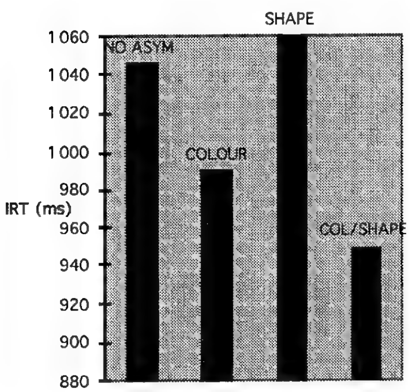


Figure 1: Mean scores of HUD condition for IRT in Experiment 1

2.3.2 Workload Data

No significant differences were found between the HUD formats for TLX Overall Workload ($F_{3,63}$=1.729, p=0.1745). Further, no significant differences were found between the conditions in any of the individual NASA TLX dimensions.

2.4 Discussion of Experiment 1 Results

Differences between the HUD formats were found for IRT only. The IRT data showed that greater asymmetry between the positive and negative pitch bars can provide a performance benefit. This reaction time difference was not supported by the other task-specific measurments. Further, no significant differences between the symbology formats were found in the workload data. Therefore, although there was a performance benefit this cannot be accounted for in terms of workload differences between the symbology formats. This may because the task-specific performance difference was in the scale of psycho-motor milli-second reaction time advantage. In this experimental task, workload may not be sensitive to such subtle differences. If this explanation is correct, it may be more appropriate to select design drivers that are of greater relevance to task performance. The milli-second performance is a consequence of decisions based upon an understanding of the situation which are in turn based upon the information displayed (i.e. the symbology). Alternatively, these data may indicate that workload an insufficient design driver to consider in this type of task. Therefore it may be more useful to attempt to understand how this decision was reached from the information provided. Workload may, in certain evaluation situations, be an inadequate design driver as it does not take into account whether the operator's knowledge state matches the requirements of the task being performed[7]. SA could be more useful in evaluation flight tasks that include spatial disorientation as they provide an indication of the observers' understanding of the aircraft attitude rather than inference of this from workload scores.

3. EVALUATION EXPERIMENT 2

3.1 Situational Awareness as a Design Driver

The rationale of using SA as design driver is that the better the symbology, the greater SA it provides the pilot, and the better pilot performance will be. Dominguez[8] examined fifteen separate definitions of SA and provided the following consolidated definition: SA is the "continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events".

Definitions of SA have commonality in that they all include the knowledge state of the operator rather than simply the attentional demands being faced. This is an important difference from workload definitions, because the way in which an individual will respond to a given level of workload will be mediated by understanding, experience, and knowledge of factors involved in the task. Thus, a pilot on a first mission, where the situations being encountered are novel, will react differently (to the same degree of workload) from an experienced pilot who has encountered the situation before and has appropriate skills and knowledge structures for dealing with it. Thus SA provides a much broader concept for describing the aircrew task, and as such provides a more useful model of human behaviour and information requirements than workload. It should be noted, however, that the two concepts are interrelated[9]. No matter how good the knowledge of a pilot, workload can always be increased to the point where overload is reached and SA lost. Therefore, any consideration of situational awareness in display design and evaluation should incorporate workload.

A number of SA measures have been developed. Fracker[10] distinguishes between (a) explicit and implicit and (b) direct and comparative measures of SA. Explicit measures require respondents to self-report material in memory of which they are consciously aware. Thus, explicit measures are deemed to be subjective. Examples of explicit measures include retrospective event recall, such as questionnaires. Implicit measures are derived from task performance and are considered objective in nature. Direct measures require a numerical value is assigned to the operator's SA in a given mission scenario. Finally, comparative measures involve comparing a pilot's SA during one mission against his SA in another and a value is assigned to the ratio of one against the other.

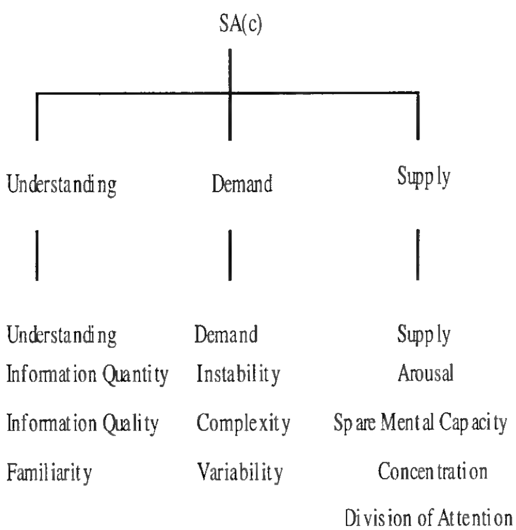


Figure 2: The Dimensions of the Situational Awareness Rating Technique

Following Fracker's definitions, SA measures have tended to rely upon an explicit approach. One such measure of SA, Situational Awareness Rating Technique[11] (SART) was developed as an aircrew systems design evaluation tool. SART provides subjective estimates of attentional demand, supply and understanding. These three constituents are postulated to be the three primary dimensions of SA. These are shown in Figure 2. Selcon and Taylor[12] conducted a study using a simulated HUD to investigate whether the three components of SART were sensitive to the performance of skill-based and rule-based tasks. It was found that the Demand component of SART was sensitive to skill-based task performance, and the Understanding component was sensitive to rule-based task performance. In addition, the same construct groupings kept emerging over several experimental sessions, further providing strong support for the internal structure of SART[12].

3.2 Method

Twelve experienced air defence pilots completed the fixed-based simulation experiment. All were volunteers with normal or corrected to normal vision.

The experiment was a within-subjects design. Each subject completed 216 trials consisting of three conditions. The positive and negative pitch bars were manipulated in colour each condition. The pitch bars in the monochrome condition were always green. In the blue/brown asymmetry condition the positive pitch bars were coloured blue and the negative pitch bars were brown. Finally, the yellow condition's HUD consisted of green positive pitch bars and yellow negative pitch bars. The conditions are summarised in Table 2.

Condition	*Positive pitch bar colour*	*Negative pitch bar colour*
No asymmetry	Green	Green
Blue/brown asymmetry	Blue	Brown
Yellow asymmetry	Green	Yellow

Table 2: Conditions of Experiment 1.

The subjects were required to recover from unusual positions (UPs) as quickly and accurately (i.e. avoiding ground-collision) as possible. The 24 unusual positions were a combination of four roll angles (45°, 135°, 225° and 315°) and six pitch angles (70°, 50°, 30°, -30°, -50° and -70°). The order of the conditions was counterbalanced across subjects using a Latin Square, and the order of presentation of UPs was randomised.

The objective measures taken were IRT (ms), TRT (ms), the number of crashes and the number of times each subject failed to recover within 15 seconds. Each subject also completed the 14-D SART scale immediately after each condition.

The pilots were allowed to practise the task until they felt comfortable with the recovery procedure and the format of the alternative HUDs. In the experimental stage, UPs were presented on a HUD display on a white background. Each pilot completed all three conditions, providing SART ratings after each condition.

3.3 Results

3.3.1 Task-Specific Performance Data

The IRT, TRT, crash and failure to recover data were analysed using balanced analyses of variance (ANOVAs) with the following fixed factors: condition, pitch angle, roll angle, order and repetition. All significant effects were tested post-hoc using the Newman-Keuls procedure. Significant differences were found between the HUD formats for IRT only. The task performance measures TRT, the number of crashes and

failures to recover did not demonstrate differences between the conditions.

Subjects' performance between HUD condition differed in terms of their IRTs ($F_{2,16}$=19.82, p<0.001). This effect can be seen in Figure 3. Subjects' IRTs were found to be significantly faster (p<0.001) under the blue/brown asymmetry condition than under both the yellow asymmetry and the no asymmetry conditions.

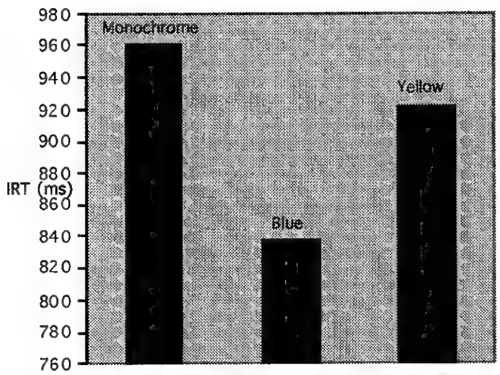


Figure 3: Mean scores of HUD condition for IRT in Experiment 2

3.3.2 Situational Awareness Data

The SART data was analysed using balanced ANOVAs. All significant effects were tested post-hoc using t-Tests adjusted using the Bonneferroni inequality to produce an experiment wise error of less than 5%. An overall measure of SA was calculated from the SART score using the formula

$$SA(c) = \Sigma U/n_U - (\Sigma D/n_D - \Sigma S/n_S)$$

where $\Sigma U/n_U$ is the mean of the scores on the Understanding related SART dimensions, $\Sigma D/n_D$ is the mean of the scores on the Demand related SART dimensions, and $\Sigma S/n_S$ is the mean of the scores on the Supply related SART dimensions.

Overall SA or SA(c) scores differed significantly between conditions ($F_{2,13}$=4.409, p<0.0264). This effect can be seen in Figure 4. The blue/brown asymmetry condition was rated as providing more overall SA than the no asymmetry condition (p<0.01). There were no other significant differences.

Further analysis of the dimensions within SA(c) showed that subjects' ratings of the symbology formats did not differ in the Demand-mean and Supply-mean dimensions. However, significant differences between the symbology formats were found for Understanding-mean scores ($F_{2,13}$=12.658, p<0.001). This effect can be seen in Figure 5. The blue/brown asymmetry condition was rated as providing significantly greater understanding than the yellow asymmetry (p<0.01) and the no asymmetry condition (p<0.0001). This implies that the differences in the SA(c) scores were mainly a product of differences in understanding, rather than attentional demand or supply.

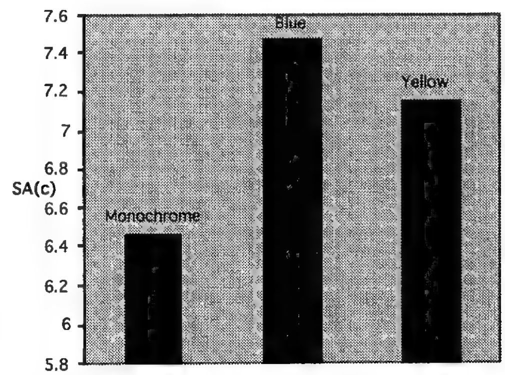


Figure 4: Mean scores of HUD condition for SA(c) in Experiment 2

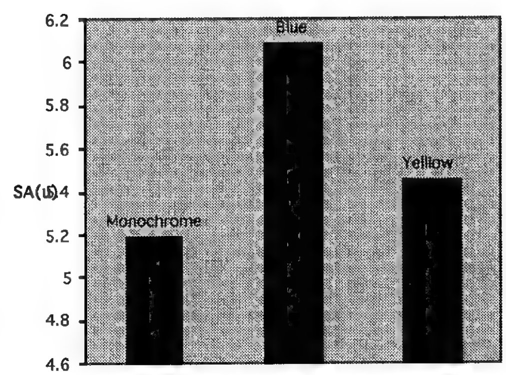


Figure 5: Mean scores of HUD condition for SA(u) in Experiment 2

Further analysis within the Understanding dimensions showed significant differences. Subjects rated the symbology formats differently in terms of information quantity ($F_{2,13}$=4.9, p<0.05). The blue/brown symbology was rated as providing greater information quantity (p<0.01) than the no asymmetry conditions. Subjects also rated the symbology formats differently in terms of information quality ($F_{2,13}$=.4.9, p<0.001). The blue/brown asymmetry symbology was rated as providing greater information quality than yellow asymmetry (p<0.05) and no asymmetry (p<0.001). This in turn implies that the differences in the Understanding-mean scores were mainly a product of differences in these two dimensions.

A significant correlation between the IRT data and the SA(c) data was found (r=0.351, p<0.05).

3.4 Discussion of Experiment 2 Results
Again, differences between symbology formats were found only for IRT. Subjects' IRTs were significantly faster under the blue/brown asymmetry condition than under both the yellow asymmetry and the no asymmetry conditions. This task-specific data was supported by the SA ratings. The blue/brown asymmetry condition was rated as providing more overall SA (SA(c)) than the no asymmetry condition. Further analysis of the SART ratings revealed that the SA differences were significant in the understanding-mean scores and not the demand- and supply-mean scores. Examination of the understanding dimension data revealed that the blue/brown symbology was rated as providing greater information quantity than the than the no asymmetry conditions and providing greater information quality than yellow asymmetry and no asymmetry. The significant SART differences between conditions are shown in *italics* in Figure 6.

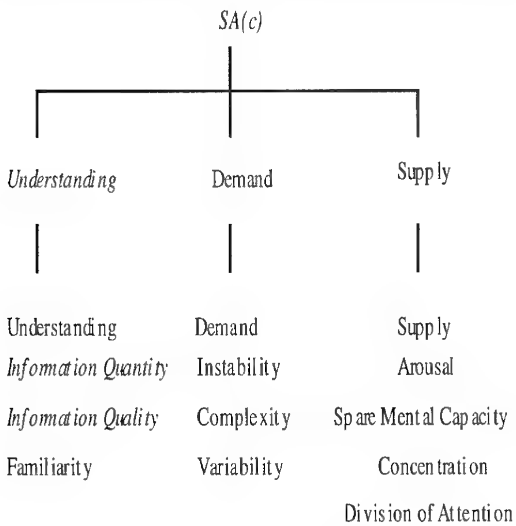


Figure 6: The Significant Dimensions of SA(c) in Experiment 2

The finding that the understanding dimension provides support the task-specific performance data has been shown in previous SART research. As described in the introduction Selcon and Taylor[11] found that the Demand component of SART was sensitive to skill-based task performance, and the Understanding component was sensitive to rule-based task performance. Recovery from unusual positions is an excellent example of a rule-based task.

4. GENERAL DISCUSSION

Significant task-specific performance advantages were found for alternative symbology formats in both the above experiments. However, although these significant differences were supported by the SA ratings they were not reflected in the workload ratings. In Experiment 1, the IRT data showed that greater asymmetry between the positive and negative pitch bars provides a reaction-time performance benefit. Similarly, in Experiment 2 subjects' IRTs were found to be significantly faster under the blue/brown asymmetry condition than under both the yellow asymmetry and the no asymmetry conditions. This performance benefit was supported and correlated with the SA ratings.

These findings suggest that SA may be a more suitable dimension to measure than workload in this experimental situation. Workload is based upon attentional demands rather than cognitive aspects of pilot performance such as prior knowledge and understanding. Indeed, in the second experiment, the attentional demand dimension of the SART scale (Supply and Demand) did not show any significant differences between conditions. Differences found between the conditions in terms of the Understanding dimension (information quantity and quality) demonstrate how design drivers such as SA can provide specific guidance to symbology designers concerning the exact nature of differences between symbology formats. This may prove to be more useful than inferring this from workload ratings in trials especially where psycho-motor performance is the result or consequence of decisions based upon an understanding of the situation which are in turn based upon the information displayed. It is possible that workload be a more useful measurement to take in evaluation situations which require multi-task performance. The boundaries concerning the application of workload or SA in evaluation trials need to be established.

The SA ratings may have been more consistent with the reaction time data due to the subject groups. The SA ratings were made by RAF pilots and the workload ratings provided by naive subjects. Differences between the conditions in terms of workload may have been limited as the subjects were inexperienced with flight tasks. In other words, because of the low skill level of the subjects, subtle differences in symbology may not effect their perception of workload. In this case, objective, rather than subjective workload measurements may be more appropriate to the task, such as secondary-task performance measurement.

Future research should consider the development of SA measures. These should consider objective as well as subjective measures. Such an alternative SA assessment method is that of Endsley[13], namely the Situation Awareness Global Assessment Technique (SAGAT). This technique takes the form of a comparison between the real situation of a simulated mission scenario and the pilot's perception of the scenario. Secondly, future research should consider the development of further design drivers. Symbology design and evaluation would benefit from the generation from a number of standardised guidelines or rules that could be measured in trials. This may be achieved through the development of the understanding dimension of the SART scale. This would involve the extension of dimensions such as understanding,

information quantity, information quality and familiarity. Further areas for the development are the dimensions within SA such as spatial and tactical awareness. This could be extended to a formal model of SA. Beyond SA, cognitive compatibility may prove to be a valid design driver. Cognitive compatibility is the design of symbology to reduce the mental processing or decoding required to use the display.

5. CONCLUSION

The results of the two symbology evaluation trials suggest that SA may a more suitable dimension than workload to measure in trials where psycho-motor task performance is critical and are the consequence of decisions based upon an understanding of the situation which are in turn based upon the information displayed (i.e. the symbology). This is because workload is based upon attentional capacity rather than cognitive aspects of pilot performance such as prior knowledge and understanding. However, workload data may be more appropriate than SA measurement in other evaluation situations, such as multi-task situations. Further research should consider where SA measurement provides the most benefit to designers and how SA can be developed in both theory and its practical application in evaluation trials. Finally, other design drivers need to be suggested that augment, but may not necessarily exclude, concepts such as workload and SA.

6. REFERENCES

1. Weinstein, L.F., Ercoline, W.R., McKenzie, I., Bitton, D.F. and Gillingham, K.K., "Standardisation of Aircraft Control and Performance Symbology on the USAF Head-Up Display", Armstrong Laboratory Technical Report NO.AL/CF-TR-1993-0088, 1993.

2. Dudfield, H.J., "A Review of the Benefits of Colour Coding Collimated Cockpit Displays", *in* "Proceedings of 21st Conference of the Western European Association of Aviation Psychologists", 1994.

3. Dudfield, H.J., Davy, E., Hardiman, T.D. and Smith, F.J., "The Effectiveness of Colouring Collimated Displays: An Experimental Evaluation of Performance Benefits", *in* "Proceedings of The 8th International Symposium on Aviation Psychology", (in press).

4. Hardiman, T.D., Dudfield, H.J., Newman, J.M., Doyle, A.J.R. and Fearnside, P.J.. "A Review of HUD and HMD Symbology. Methodology for Design and Evaluation.", Unpublished MoD Report, October 1994.

5. Taylor, R.M. and Selcon, S.J., "Understanding Situational Awareness, Contemporary Ergonomics, 1990, pp 105-110.

6. Kantowitz, B.H. and Casper, P.A., Human Workload in Aviation" *in* Wiener, E. and Nagel, D. (Eds.), "Human Factors in Aviation", San Diego, California, Academic Press, 1988, pp 150-171.

7. Selcon,S.J., Coxell, A.W., Bunting, A., Smith F.J., Irving, M.A., Lal, R., Dudfield, H.J. and Shorter, N., "The Development of an Explanatory Tool for Image-Fused Displays", Unpublished MoD Report, February 1995.

8. Vidulich, M., Dominguez, C., Vogel, E. and McMillan, G., "Situtation Awareness: Papers and Annotated Bibliography", Armstrong Laboratories, Al-TR-19940085, 1994.

9. Selcon, S.J., Taylor, R.M. and Koritsas E., "Workload or Situational Awareness? TLX vs. SART for Aerospace System Design Evaluation", *in* "Proceedings of the Human Factors Society 35th Annual Meeting", Human factors Society, Santa Monica, 1991.

10. Fracker, M.L., "Measures of Situation Awareness: Review and Future Directions", Armstrong Laboratories, 1991.

11. Taylor, R.M., "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design", *in* "Proceedings of Situational Awareness in Aerospace Operations", AGARD CP-478, 1990, Paper 3.

12. Selcon, S.J. and Taylor, R.M., "Evaluation of the Situational Awareness Rating Scale (SART) as a Tool for Aircrew Systems Design", *in* "Proceedings of Situational Awareness in Aerospace Operations", CP-478, 1990, Paper 5.

13. Endsley, M.R., "A Methodology for the Objective Measurement of Pilot Situation Awareness", *in* "Proceedings of Situational Awareness in Aerospace Operations", CP-478, 1990, Paper 1.

Scene-linked Symbology to Improve Situation Awareness

Robert S. McCann & David C. Foyle

Flight Management and Human Factors Division (MS 262-3)
NASA Ames Research Center
Moffett Field, CA 94035-1000
USA

1. SUMMARY

This paper reviews recent research conducted in the Flight Management and Human Factors Division of NASA Ames Research Center on superimposed symbology (as found on HUDs and HMDs). We first identify various performance problems which suggest that superimposed symbology impairs pilots' ability to maintain simultaneous awareness of instrument information and information in the forward visual scene. Results of experiments supporting an attentional account of the impairment are reported. A design solution involving the concept of "scene-linked" symbology is developed, and experiments testing the design solution are reported. An application of the scene-linking concept, in the form of a candidate HUD to support ground taxi operations for civil transport, is described.

2. INTRODUCTION

2.1. Information acquisition and situation awareness

Piloting an aircraft is a demanding activity, in part because pilots have to be aware of many different forms of information. Most of this information is extracted from two distinct sources: the instrument panel, or "near domain", and the forward visual scene, or "far domain". In a standard flight deck, the instrument panel is located underneath the windshield, making it physically impossible to see both domains simultaneously. As a result, pilots must adopt a sequential acquisition scanning strategy whereby information is sampled from one domain and then the other. This strategy requires time consuming actions, such as eye and head movements, and reaccommodation of the eyes. These actions continually interrupt the process of information acquisition. Furthermore, as long as the pilot is looking at one domain, a sudden event (or sudden state change) in the other domain goes undetected.

For various reasons, then, the physical separation between near and far domains has a negative effect on situation awareness. At first glance, the problem would seem to be solved by Head-Up Displays (HUDs) and Helmet-Mounted Displays (HMDs), which superimpose graphic depictions of instrument symbology directly over the far domain. By bringing near and far domains into the same forward field of view, superimposed symbology devices make it physically possible to process near and far domains in parallel (Ref 1, 2). Intuitively, then, it would seem that these devices would enhance a pilot's situation awareness, relative to the traditional configuration.

2.2. Performance problems

Over the years, however, researchers have identified a number of performance problems with superimposed symbology. These problems suggest that, far from facilitating joint awareness of near and far domains, superimposed symbology actually reduces the level of joint awareness. For example, Fischer, Haines, & Price (Ref 3) found that pilots flying simulated approaches using a HUD sometimes failed to notice runway incursions. No such failures were observed among pilots flying with conventional head-down instrumentation. Weintraub, Haines, & Randle (Ref 4) found similar results using static displays. Fischer et al. confounded location of the instrumentation (superimposed versus head down) with type of instrumentation; the HUD included contact analog symbology, whereas the head-down instrumentation did not. It is not clear, then, whether the failure to notice incursions was due to the change in the location of the symbology or to the change in the symbology itself. Wickens & Long (Ref 5) recently addressed this problem by presenting the identical symbol set either head-down or head-up. Following breakout, pilots flying instrument approaches took, on average, 2.5 seconds longer to respond to an unexpected runway incursion when the symbology was head-up compared to head-down.

A second performance problem has emerged from level flight simulation tasks at NASA-Ames. Brickner (Ref 6) had subjects fly a simulated helicopter through a slalom course demarcated by virtual pylons. Subject pilots were instructed to fly around the pylons while maintaining an altitude of 100 feet. In one condition, altitude information was available only from naturally occurring environmental cues in the graphic simulation of the far domain (e.g., pylon size). In another

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

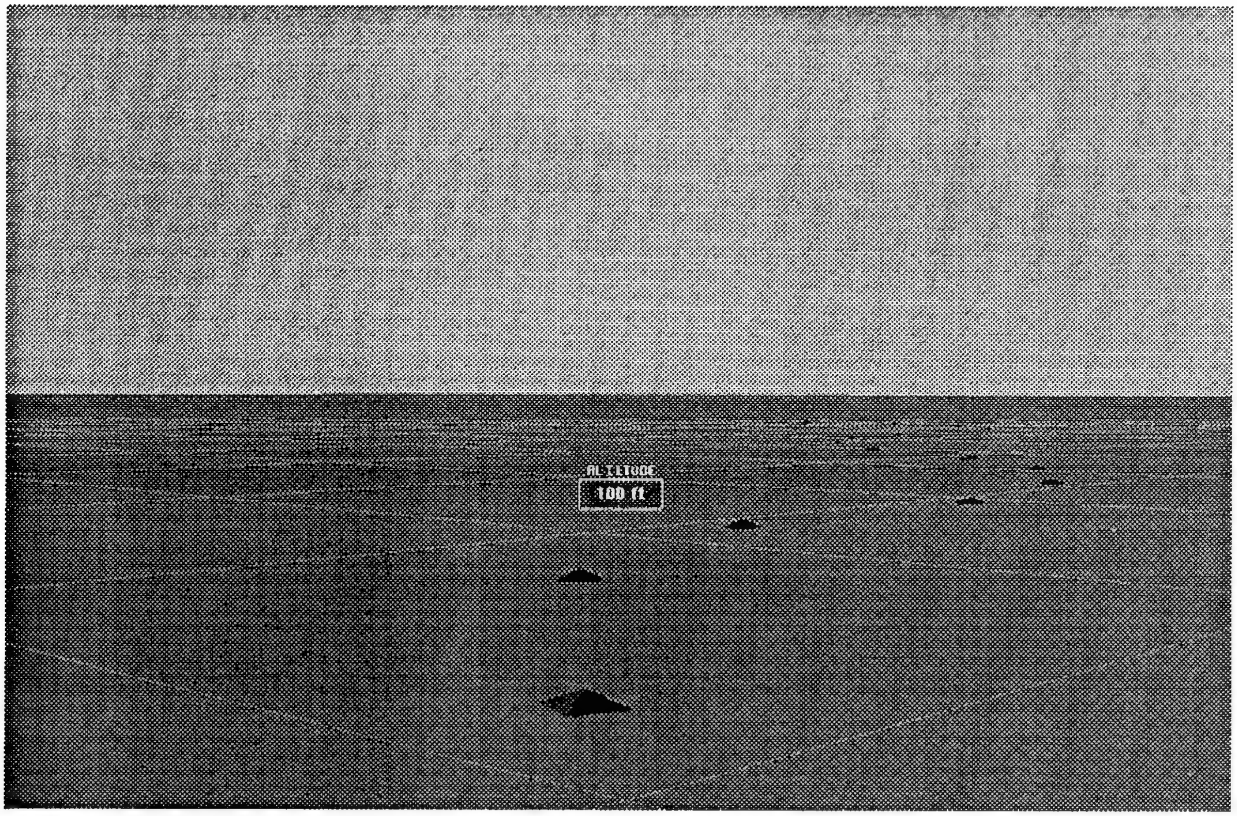


Figure 1. Part-task simulation environment showing ground track to be followed (pyramids) and digital superimposed symbology (currently showing 100 ft). (After Foyle et al., Ref 7).

condition, these natural cues were supplemented by a superimposed digital readout of current altitude (for brevity, we refer to the digital symbol as a "HUD"). Not surprisingly, the presence of the digital HUD improved altitude maintenance performance compared to the no-HUD condition. However, this performance benefit was obtained at the cost of an increase in the number of collisions with the pylons. Thus, superimposing digital symbology on the forward visual scene yielded a performance tradeoff: the symbology supported more accurate altitude maintenance, at the cost of less accurate path maintenance.

Foyle, McCann, Sanford, & Schwirzke (Ref 7) found a similar tradeoff using a slightly different flight task and a different performance measure. Subjects flew a curving path defined by small pyramids on the ground, while maintaining an altitude of 100 feet (see Figure 1). Random buffeting was introduced in both the vertical and horizontal dimensions throughout each 2-minute flight; the dependent measures were flight path error (measured by root mean square deviations from the designated path) and altitude error (measured by root mean square deviations from 100 feet). Following Brickner (Ref. 6), altitude information was available either from environmental cues in the simulated far domain, or environmental cues supplemented by a digital altitude HUD. Results were similar to Brickner's: the presence of the HUD decreased altitude maintenance error, but increased path following error. In subsequent discussion, we refer to this performance pattern as the altitude/path performance tradeoff.

2.3. Source of the performance problems

These performance problems suggest that superimposed symbology actually reduces a pilot's joint awareness of events in the near and far domains. Why is this the case, when common sense indicates that superimposing symbology on the far domain should have the opposite effect? One possibility, discussed by Roscoe and his colleagues (Ref 8), is that even though superimposed symbology is collimated to appear at visual infinity, there are still a variety of perceptual cues to remind the pilot that the symbology is much closer than the far domain (for example, scratches or dirt on the combiner glass of a HUD). Therefore, when processing superimposed symbology the eye accommodates inward, blurring the out-the-window

scene to where concurrent processing of the symbology and the world is prevented. However, this account cannot explain the altitude/path performance tradeoffs reported by Brickner (Ref 6) and Foyle et al. (Ref 7), or the increased time to notice runway incursions reported by Wickens & Long (Ref 5). In these studies, the superimposed symbology and the out-the-window scene were part of the same synthetic graphical display; thus, both the superimposed symbology and the far domain were at the same optical distance from the eye.

A second possibility appeals to limitations on the ability of the visual system to process superimposed symbology and the world simultaneously (Ref 9, 3, 7). This hypothesis follows naturally from "object-based" models of visual attention (Ref 10, 11). According to these models, visual processing occurs in two successive stages. In the first stage, visual elements with similar perceptual properties are grouped together to form distinct perceptual units (Ref 10, 12, 13). HUD symbology differs from the far domain on a number of salient dimensions, including color, texture, and motion (HUD symbology is either stationary or moves over a small visual area, whereas elements in the far domain are linked in a common flow field). Each of these dimensions is a powerful basis for perceptual grouping (Ref 10, 12, 13). Thus, in the first stage of processing, superimposed symbology is parsed as one perceptual group, and the far domain as another.

In the second stage, perceptual groups form the basis of attentional allocation. Importantly, limitations on visual attentional resources prevent attention from being focused on more than one perceptual group at any one time (Ref 11). Therefore, when superimposed symbology is selected for processing, it captures all available attention. Since elements in unattended groups are not processed to the point of awareness (Ref 14), attentional capture causes pilots to lose awareness of events or elements in the far domain.

Object-based models thus provide a natural account of the performance problems described earlier. The increased latency to respond to runway incursions when using HUDs (Ref 5) follows from the fact that when pilots are attending to the HUD, far domain awareness is reduced to the point where runway incursions are not noticed. The altitude/path performance tradeoffs reported by Brickner (Ref 6) and Foyle et al. (Ref 7) follow from the fact that, because attentional capture by the digital HUD reduces awareness of the far domain, departures from the flight path take longer to be noticed and corrected.

Attentional capture by superimposed symbology poses a challenge to operational efficiency and safety that grows more serious every day. This is because superimposed symbology devices are spreading rapidly beyond the military sector, where they have existed for many years. For example, one of the largest US. carriers, Southwest Airlines, is currently retrofitting its entire fleet with HUDs. HUDs are also now available to the general aviation market. In the near future, superimposed symbology devices are likely to be incorporated into a host of additional operating environments. These include automobiles, industrial assembly lines, and occupations, such as fire fighting, where people must operate in low-visibility conditions.

Here at NASA Ames, concerns about the operational implications of attentional capture have motivated two lines of research. One line has verified a key empirical prediction of the capture account, and identified the perceptual characteristic most responsible for capture. The other line has incorporated this information into candidate HUD displays, which are then tested to determine whether they alleviate the performance problems associated with capture. The rest of this article summarizes these programs.

3. TESTS OF ATTENTIONAL CAPTURE

3.1. Introduction

Consider the following situation. A pilot is viewing a visual scene consisting of a runway, just prior to touchdown, together with superimposed symbology on a HUD. The task is simply to identify two discrete objects. If superimposed symbology captures attention, processing the two objects should proceed in parallel if they are both HUD symbols. This follows from the fact that the symbols are part of the same perceptual group, and attention is distributed equally across the elements of a perceptual group (Ref 10). However, if one object is a HUD symbol and the other is a feature on the surface of the runway, processing the two objects should be serial, because attention must be switched from the HUD to the far domain before the object on the runway can reach awareness. Since switching attention takes time, the attentional capture hypothesis makes a straightforward prediction: the pilot should respond to the task more slowly when one object is a HUD symbol and the other object is an element on the runway surface, compared to when both objects are HUD symbols. We recently completed a series of laboratory experiments testing this prediction (15, 16, 17).

3.2. Experiment 1: A Test of Attentional Capture

Following Weintraub et al. (Ref 4), subjects viewed computer-generated displays consisting of a set of stationary blue symbols (collectively referred to as the HUD) superimposed on a yellow image of a runway (see Figure 2). All far domain imagery (including the runway outline, surface features on the runway, and the

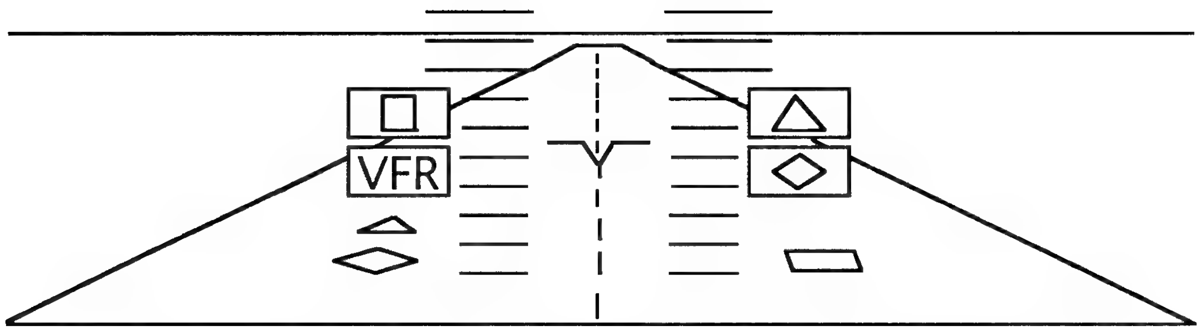


Figure 2. HUD Symbology superimposed on runway scene. Subjects' task shown was to identify VFR cue (on HUD), then visually acquire diamond (lower left on runway). (After McCann et al., Ref 15).

horizon line) was dynamic, consistent with the appearance of the far domain during final approach. The task was to first identify a three-letter cue located on the HUD. Depending on cue identity, subjects then searched either the remaining HUD symbols or the symbols on the runway surface for one of two prespecified targets - a stop sign or a diamond. If subjects saw a diamond, the runway was open, meaning the landing could continue. If subjects saw a stop sign, the runway was closed, meaning a go-around was mandated. The decision to land or go around was communicated by pressing one button if the target was a diamond, and another button if the target was a stop sign. Instructions stressed the importance of responding to the task both quickly and accurately.

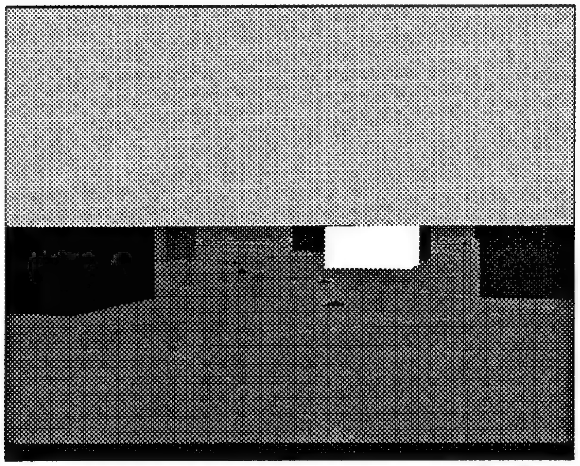
Importantly, the targets and distracting symbols were the same on the HUD and on the runway; the cue signalled which domain was relevant to search, and which was not. Nevertheless, responses were approximately 100 msec faster when the relevant target was on the HUD compared to the runway surface. This result was not because the runway versions of the targets were inherently more difficult to process than the HUD versions. When the cue was altered to make it look like it, too, was on the runway surface, the response time pattern reversed: subjects were now slower when the relevant target was on the HUD than when it was on the runway. And since the displays also equated the physical distance between cue and targets between and across perceptual groups, attention switching between the HUD and the far domain provided the most straightforward account of the data. Thus, the results fully supported the hypothesis that well defined perceptual groups, such as superimposed symbology on a HUD, capture attention.

3.3. Experiment 2: What causes capture?

According to object-based models of attention, capture occurs because the visual system parses superimposed symbology as one perceptual group, and the far domain as another. In the course of developing a design solution to the problem, our first step was to identify the perceptual characteristic, or combination of characteristics, most responsible for perceptual grouping. In the first experiment, superimposed symbology was distinguished from the far domain by a number of highly salient characteristics, including differential motion, differential color, and differential viewing perspective (the HUD symbology was vertical with respect to the viewer, whereas objects in the far domain appeared as they would when viewed from above and behind). Which of these characteristics was most important in driving perceptual grouping, and hence attentional capture?

McCann et al. (Ref 16) examined the contributions of differential color and differential motion to the grouping effects in McCann et al (Ref 15). A baseline condition was provided by replicating McCann et al. (Ref 15), where the HUD symbology was distinguished from the far domain by differential motion, color, and viewing perspective. The remaining conditions were created by jointly manipulating whether the HUD and the world were shown in the same or different colors, and whether the point of regard with respect to the runway was dynamic, consistent with final approach, or "frozen" at about 5 seconds prior to touchdown. Since the HUD and the elements of the far domain were both stationary in this condition, there were no differential motion cues to support grouping.

The logic of these manipulations can be illustrated with reference to the color factor. If perceptual grouping is driven by color differences between the HUD and the far domain, then parsing the HUD and the far domain as separate perceptual groups should not occur when the HUD symbology and the far domain are drawn in the same color. In the absence of separate grouping, there should be no attentional capture. Processing should be the same regardless of whether the cue and target are both superimposed symbols, or the cue is part of the HUD and the target on the runway. Empirically, response times should be the same across the two



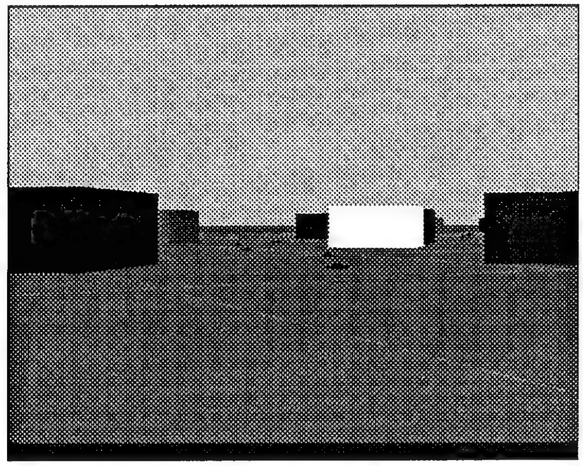


Figure 3. Flight simulation environment with virtual buildings showing current altitude at 100 feet (top panel), and below 100 ft (bottom).

conditions. Alternatively, if color is not a factor in the grouping process, the HUD should continue to be parsed as a perceptual group, distinct from the far domain. Response times should continue to be slower when the target is on the runway surface compared to the HUD. In general, our interest is in comparing the difference in response times across the two critical conditions (cue and target on the HUD versus cue on the HUD and target on the runway) to the difference obtained in the baseline condition. We can then determine whether attentional capture is driven primarily by differences between superimposed symbology and the far domain in color, in motion characteristics, or another characteristic entirely (such as viewing perspective).

The results of the baseline condition replicated the earlier finding (Ref 15) that responses were slower when the cue was on the HUD and the target was on the runway surface, compared to when cue and target were both on the HUD. This difference was virtually unchanged when the superimposed symbology and the far domain were presented in the same color. In sharp contrast, when differential motion cues were removed from the display, the difference in response time between the two critical conditions was reduced by 50 percent, a highly significant effect.

3.3.1. *Implications*

The purpose of this experiment was to identify which of the perceptual characteristics distinguishing superimposed symbology from the far domain was most responsible for perceptual grouping (and hence, attentional capture). Although differential color was an obvious candidate, the experiment suggests that differential motion, not color, plays an important role. These results have direct implications for display design. If color had been found to cause attentional capture, capture could have been reduced by simply drawing HUD symbology in colors that match the far domain. Clearly a more complex design solution is required. One possibility is considered in the next section.

4. A CANDIDATE DESIGN SOLUTION

If the primary driver behind attentional capture is differential motion between superimposed symbology and the far domain, then capture should be prevented if differential motion between the HUD symbology and the elements of the far domain is removed. A design option that achieves this goal involves replacing conventional HUD symbols with virtual symbols that appear to be physically part of the world (Foyle, Ahumada, Larimer, & Sweet, Ref 18). As the aircraft moves through the world, these "scene-linked" symbols undergo the same visual transformations as real objects. There are no differential motion cues to cause the visual system to interpret the virtual symbols as part of a perceptual group distinct from the world. In the absence of such parsing, attentional capture should be prevented, enabling pilots to process scene-linked HUD symbology in parallel with information in the far domain.

5. EXPERIMENTAL TESTS

5.1. Experiment 1: Virtual Buildings

If this analysis is correct, scene-linked symbology should alleviate performance problems found with conventional forms of superimposed symbology. A

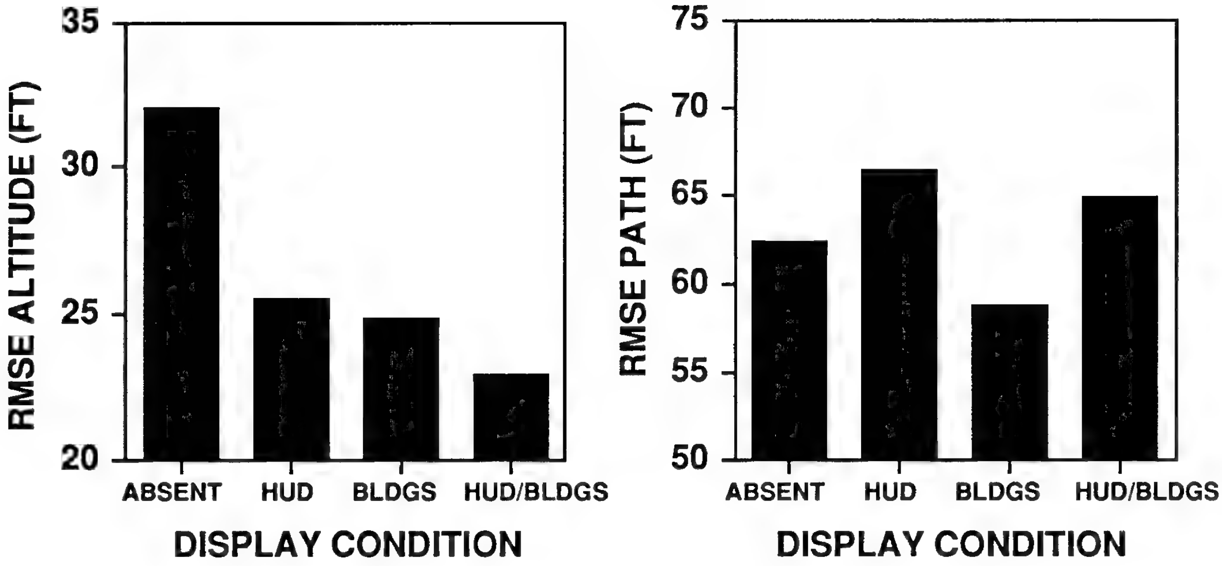


Figure 4. Effects of HUD altitude symbology absence, presence, virtual buildings, and virtual buildings with altitude symbology on RMS Error Altitude (left) and RMS Error Path (right).

recent experiment examined the effect of scene-linking on Foyle et al.'s (7) altitude/path performance tradeoff (the finding that in a part-task simulation of helicopter flight, superimposing a digital altitude indicator improved altitude maintenance performance, but impaired path following performance). In addition to the standard condition involving the superimposed digital altitude symbol, we included a condition in which "virtual" buildings were added to both sides of the path at regular intervals. Each building was exactly 100 feet in height, the assigned maintenance altitude. The two panels in Figure 3 illustrate the various cues to altitude supplied by the buildings. In the left panel, the vehicle is at 100 feet, and is flush with the tops of the buildings. Additionally, as determined by the visual geometry, the tops of the buildings are coincident with the horizon line. In the right panel the vehicle is below 100 feet, so the buildings now extend above the horizon. Thus, the buildings provide a number of high quality visual cues to altitude.

5.1.1. *Results*

The results are presented in Figure 4. The left panel shows that, as expected, the presence of digital altitude information improved altitude maintenance relative to the control condition. The virtual buildings also improved altitude maintenance, by an amount equal to the digital HUD. The right panel shows that, relative to the control condition, the digital HUD yielded a decrement in path performance, replicating the altitude/path performance tradeoff found in earlier work (Ref 7). However, there was no decrement in path performance with the virtual buildings. The digital HUD was associated with an altitude/path performance trade-off, but the scene-linked HUD was not.

5.1.2. *Discussion*

These results demonstrate that scene-linked symbology can be just as effective as traditional forms of superimposed symbology when it comes to providing information. This follows from the fact that the improvement in altitude maintenance associated with the virtual buildings was equal to the improvement associated with the digital HUD. Unlike the digital HUD, however, the virtual buildings did not produce a decrement in path following. At a theoretical level, this result suggests that scene-linking the altitude cues enabled concurrent processing of HUD symbology and information in the far domain. At a practical level, the result supports our contention that scene-linked HUDs provide a design solution for performance problems associated with attentional capture.

5.2. Experiment 2: Scene-linking versus ease of processing

Although the buildings experiment was informative, it left an important question unresolved. The path-following component of the flight task was based on perceived distance between the helicopter and the tops of the pyramids - an analog form of computation. Similarly, when altitude cues were provided by the

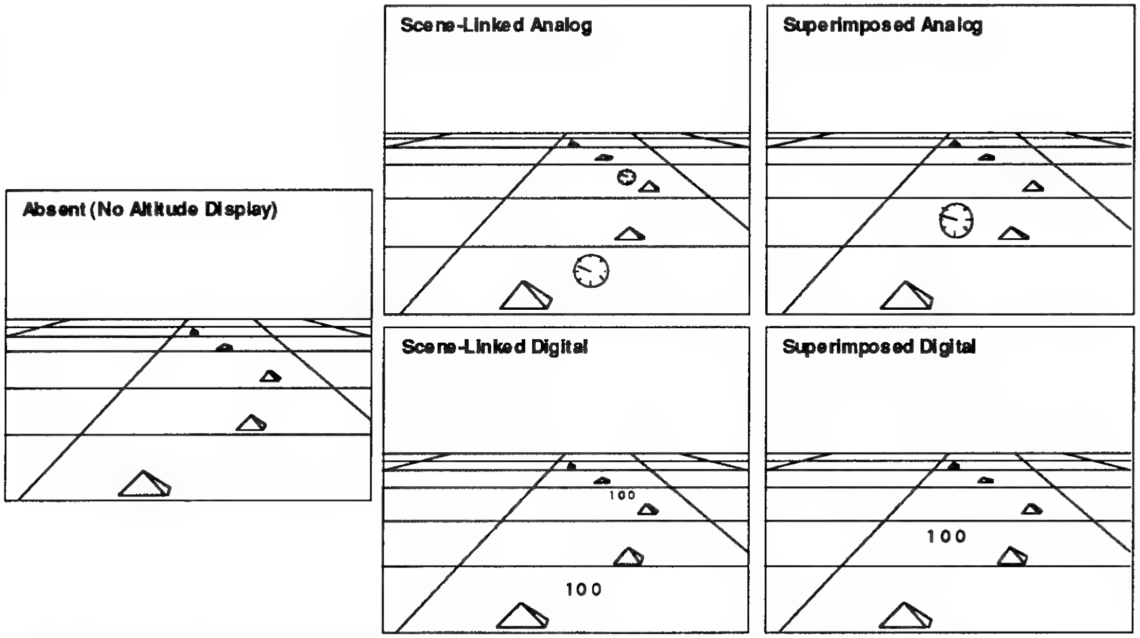


Figure 5. Schematic drawings (not to scale) of the five HUD symbology conditions (as labeled).

virtual buildings, altitude maintenance was based on the perceived distance between the vehicle and the tops of the buildings - also an analog computation. However, when the altitude cue took the form of a superimposed digital HUD, altitude maintenance was based on a digital computation. Scene-linking was thus confounded with the form of the altitude information. In general, analog displays are thought to be easier to process than digital displays; analog information is extracted more intuitively, it maps more directly onto the response system (i.e., analog control inputs), and it requires fewer mental transformations. Thus, it is not clear from the experiment whether the virtual buildings improved concurrent processing of the HUD symbology and the far domain because the buildings were scene-linked, or because they provided altitude information in a form that was easier to process than digital information.

We recently completed an experiment to discriminate the scene-linking account from the different format account (Foyle, McCann, & Shelden, Ref 19). One test involved a scene-linked version of the digital altitude indicator, where the digital readout was converted to a virtual object and interleaved with the pyramids (illustrated in Figure 5). On the one hand, if parallel processing of the superimposed digital HUD and the path was discouraged because of difficulty processing digital information, the same difficulty should be present when the digital symbology is scene-linked. Consequently, the altitude/path performance trade-off found with the superimposed symbol should be preserved. On the other hand, if parallel processing was prevented due to superimposed symbology capturing visual attention, then the scene-linked version should enable parallel processing, just as the scene-linked buildings did. Therefore, the altitude/path performance tradeoff should disappear.

The other test required an analog symbol for altitude that could be either scene-linked or superimposed. These criteria were satisfied by a "clockface" containing a pointer to current altitude (Figure 5). When the helicopter was flying at exactly 100 feet, the pointer was at the 9 o'clock position. Deviations below 100 feet caused the pointer to rotate in a counter-clockwise direction; hence, as the helicopter descended, the pointer rotated downward. Similarly, deviations above 100 feet caused the pointer to rotate clockwise, in an "up" direction. As with the digital altitude display, this analog display was presented either superimposed (Figure 5; top right panel), or as a scene-linked virtual object interleaved with the pyramids.

The predictions are straightforward. If the altitude/path performance tradeoff found in earlier studies (Ref 7) was due to greater difficulty processing digital than analog display formats, the tradeoff should be

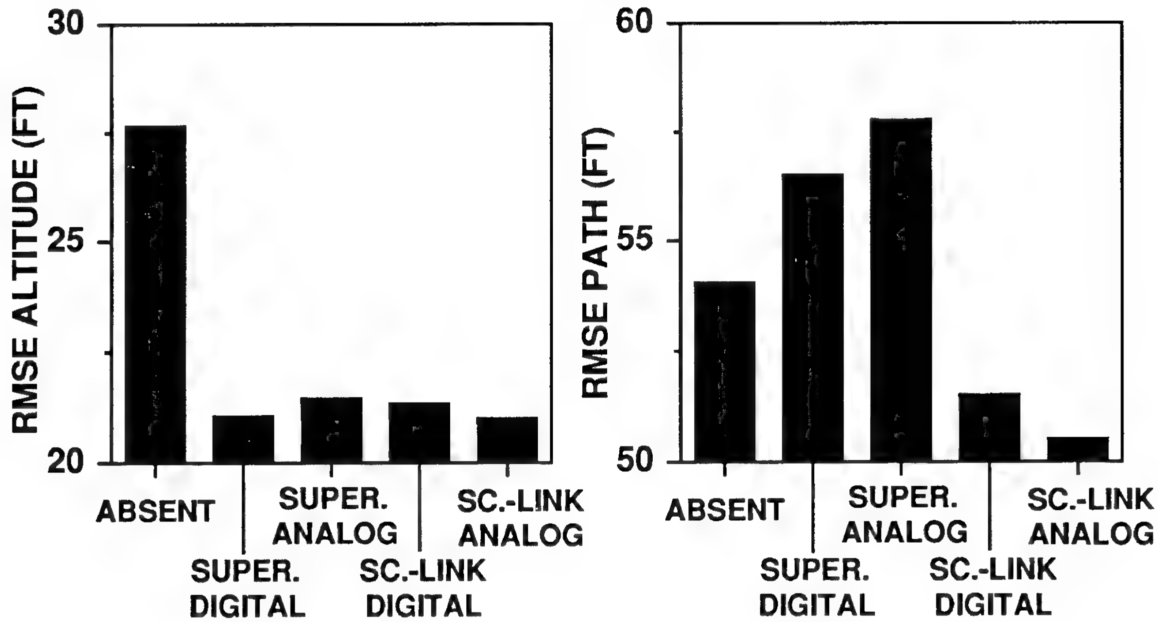


Figure 6. Results of experimental test: Effects of HUD altitude symbology absence, superimposed digital symbology, superimposed analog clock symbology, scene-linked digital symbology and scene-linked analog clock symbology on RMSE Altitude (left) and RMSE Path right).

eliminated by the clockface altitude display, regardless of whether the display is superimposed or scene-linked. Alternatively, if scene-linking is the critical factor, then the performance tradeoff should be present when the clockface is superimposed on the forward scene, but not when the clockface is scene-linked.

5.2.1. *Results*

The results are summarized in Figure 6. Starting with altitude maintenance (left panel), we see that, relative to the control condition, all of the altitude displays yielded better performance. Statistical analyses confirmed this observation, and also revealed that the magnitude of the benefit was the same for all displays. We conclude, therefore, that the clockface display was just as useful a guide to altitude as the digital display. The right panel shows that, relative to the control condition, the improvement in altitude maintenance was accompanied by an increase in path following error for the superimposed versions (both digital and analog formats). This replicates the altitude/path performance tradeoff found in previous experiments. In sharp contrast, the scene-linked displays (both analog and digital) yielded a significant *decrease* in path error.

5.2.2. *Discussion*

The results can be summarized as follows. An altitude/path performance tradeoff was present when the altitude display was superimposed on the far domain, but not when the display was scene-linked. This was true regardless of whether the form of the altitude display was digital or analog. We infer from this pattern that scene-linking produced the performance benefits obtained in the buildings experiment, not the change in display format that accompanied scene-linking.

One aspect of the results deserves additional comment. This is the fact that, relative to the control condition, the scene-linked altitude displays not only afforded an improvement in altitude maintenance, but also in path maintenance. The latter result may be due to the fact that the scene-linked displays, being interleaved with the pyramids, increased the number of reference points against which to gauge the helicopter's current position relative to the path. Regardless of the source of the benefit, it illustrates an important point. As well as promoting parallel processing of superimposed symbology and the far domain, scene linked symbology can enhance or augment flight-relevant information in the far domain. Thus, scene-linking offers not one, but two opportunities to enhance performance.

6. IMPLICATIONS AND FUTURE DIRECTIONS

Design solutions are only useful insofar as the technology is available to implement them. We should note that certain components of a scene-linked HUD are already in place, in the form of fully conformal

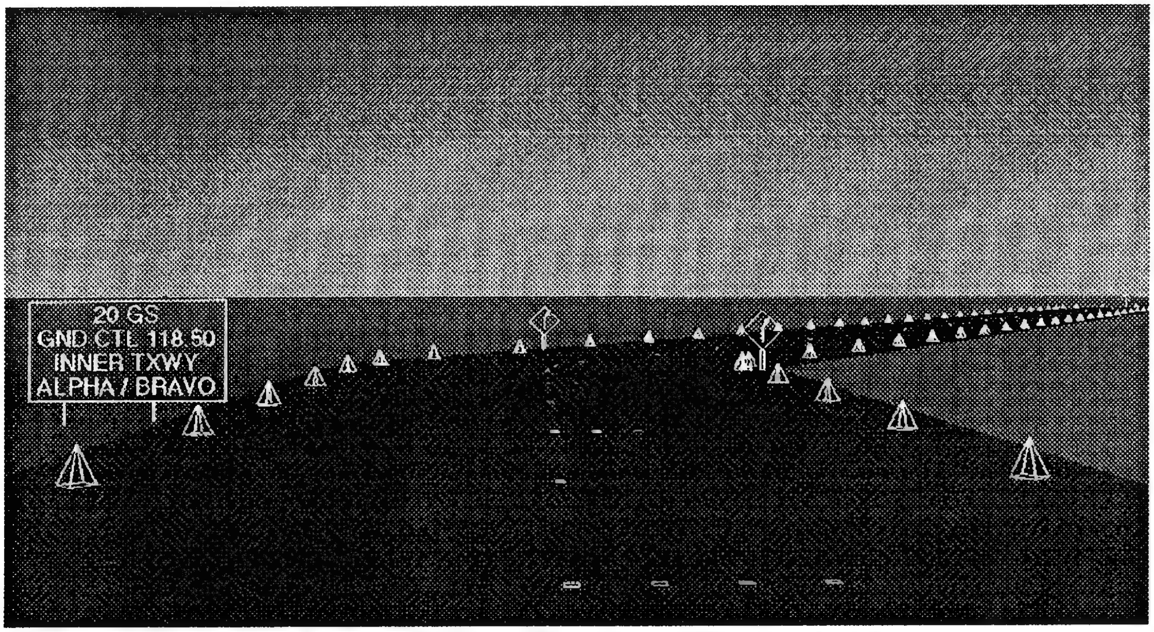


Figure 7. Scene-linked HUD symbology for taxi and surface operations. Symbology (shown in white) includes Virtual Instruments (billboard aircraft instrumentation and location information) and virtual Scene Augmentations (edge cones, turn signs and "countdown" warnings).

runway outlines. The technology necessary to generate this and other scene-linked symbology requires an advanced display media, such as a holographic HUD, a highly accurate positioning system, and a visual database. Today, positioning systems are only available at airports equipped with precision radar facilities. In the near future, however, satellite-based positioning systems (GPS) will bring accurate positioning capability to virtually all aircraft. As GPS systems saturate the marketplace, there is no technical reason why scene-linked HUDs could not preliferate along with them.

Our research suggests that scene linking superimposed symbology abolishes performance problems associated with attentional capture. If designed appropriately, scene-linking can also improve performance on tasks, such as guidance and navigation, that are based on far-domain information. These features should make scene-linked symbology particularly useful in three environments. One is nap-of-the-earth helicopter flying, where rapid switching between the instruments and the out-the-window scene is a constant requirement. Another is low visibility approaches, since pilots are focusing on primary flight display symbology, but must at the same time be sensitive to runway incursions, other air traffic, and ground traffic. The third is low visibility taxi operations. Enhancing or augmenting far domain information with scene-linked symbology could lead to faster and more efficient taxi operations, and perhaps even enable taxi operations under low visibility, where none are permitted today. The development of low-visibility scene-linked HUD symbology for airport taxi is currently underway at NASA, and is discussed below.

6.1. Scene-linked taxi symbology

Surface operations are a particularly attractive option for scene-linked HUDs. Currently, surface operations are one of the least technologically sophisticated components of the air transport system. Pilots are given little or no explicit information about their current position, and routing information is limited to ATC communications and airport charts. Under low visibility conditions, pilots can become spatially disoriented, leading to time-consuming interactions with ATC and reductions in taxi speed. Figure 7 illustrates a candidate scene-linked HUD symbology taxi display to alleviate the problems. The symbology contains two types of scene-linked information: virtual instruments (aircraft communication information and current location displayed on a virtual "billboard"), and scene augmentations (taxiway edge markers pictorially augmenting the scene).

The virtual billboard to the left of the taxiway includes aircraft communication status information and ground location. The top line contains the aircraft's current

ground speed (20 KTS, "20 GS"). This is a dynamic readout and would change as appropriate. Similarly, the ground billboard represents the aircraft's current airport location. The "Current, Last/Next" format represents current runway or taxiway segment ("Inner Taxiway"), the last intersection passed ("Alpha"), and the next intersection upcoming ("Bravo"). The example shows that this aircraft is on the Inner Taxiway, past Alpha, and before Bravo.

The pictorial scene augmentations shown include visual information that would aid the pilot in following the taxiway clearance and completing turns. Vertical side cones on the side of the commanded taxiway path depict the ATC cleared route on the HUD in superimposed symbology (as in "Pink 5" at Chicago O'Hare). The cones are conformal and represent a virtual representation of the cleared taxi route on the HUD. Both the cones and the centerline markings are shown repeated every 50 feet down the taxiway. The vertical development and constant spacing should yield increased capability for estimating ground speed, drift, and look-ahead capability for turns (see Denton, Ref 20; Johnson & Awe, Ref 21). Turn "countdown" warnings are shown in which each turn has countdown (4, 3, and 2) centerline lights that are (300, 200, and 100 feet, respectively) before each turn. This gives added distance cues for the turn. The virtual turn signs (with the arrows) give an added cue to the turn. In addition, the angle of the arrow on the sign represents the true angle of the turn (i.e., 30 deg right for a 30 deg right turn). All of the HUD symbology is scene-linked, enabling the pilot to process the symbology and still retain awareness of other traffic, including possible incursions. This and other candidate scene-linked HUDs are currently under test in a high-fidelity part-task simulator at NASA.

7. CONCLUSION

This article has reviewed recent research on superimposed symbology in the Flight Management and Human Factors Division at NASA-Ames Research Center. The message from our work cam be summarized as follows. Human information processing abilities are severely constrained by attentional limitations. These limitations must be taken into consideration when evaluating the costs or benefits of a particular display device. In the present case, we have seen that superimposing symbology on the pilot's forward field of view is necessary but not sufficient to support simultaneous processing of instrument information and far domain information. Concurrent processing can be achieved, however, with scene-linked symbology.

8. ACKNOWLEDGMENTS

We appreciate the experimental assistance of Jean M. Lynch, Beverly Sanford, Martin Schwirzke, Steve Sheldon, and Dominic Wong, and thank James C. Johnston for discussion of some of these issues. Supported by NASA RTOPS 505-64-36, 505-64-53, and 538-04-13.

9. REFERENCES

1. Naish, J. M. (1964). Combination of information in superimposed visual fields. *Nature*, *202*, 641-646.

2. Lauber, J. K., Bray, R. S., Harrison, R. L., Hemingway, J. C., & Scott, B. C. (1982). *An operational evaluation of head-up displays for civil transport operations* (NASA Technical Paper 1815). Moffett Field, CA: NASA Ames Research Center.

3. Fischer, E., Haines, R. F., & Price, T. A. (1980). *Cognitive issues in head-up displays*. NASA Technical Paper 1711, NASA Ames Research Center, Moffett Field, CA.

4. Weintraub, D. J., Haines, R. F., & Randle, R. J. (1984). The utility of head-up displays: Eye-focus versus decision time. *Proceedings of the 28th Annual Meeting of the Human Factors Society* (pp. 529-533). Santa Monica, CA: Human Factors Society.

5. Wickens, C. D. & Long, J. (1995, in press). Object vs. space-based models of visual attention: Implications for the use of head-up displays. *Journal of Experimental Psychology: Applied.*

6. Brickner, M. S. (1989). Apparent limitations of head-up displays and thermal imaging systems. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 703-707). Columbus, OH: The Ohio State University.

7. Foyle, D. C., McCann, R. S., Sanford, B. D., & Schwirzke, M. F. J. (1993). Attentional effects with superimposed symbology: Implications for head-up displays (HUD). *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1340-1344). Santa Monica, CA: Human Factors and Ergonomics Society.

8. Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. (1988). Eye accommodation to head-up virtual images. *Human Factors*, *30*, 689-702.

9. Fischer, E. (1979). *The role of cognitive switching in head-up displays*. NASA Contract Report 3137). Moffett Field, CA: NASA Ames Research Center.

10. Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual Organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.

11. Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501-517.

12. McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 55-64.

13. Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295-340.

14. Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 275-285.

15. McCann, R. S., Foyle, D. C., & Johnston, J. C. (1993). Attentional limitations with Head-Up Displays. In R. S. Jensen (Ed.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 70-75). Columbus, OH: The Ohio State University.

16. McCann, R. S., Lynch, J. M., Foyle, D. C., & Johnston, J. C. (1993). Modeling attentional effects with Head-Up Displays. *Proceedings of the 37th Annual Meeting of the Human Factors Society* (pp. 1345-1349). Santa Monica, CA: Human Factors Society.

17. McCann, R. S., Johnston, J. C., Foyle, D. C., & Lynch, J. M. (1993, November). *Shifting attention between visual domains: Evidence from Head-Up Displays*. Paper presented at the 34th annual meeting of the Psychonomics Society, Washington, D. C.

18. Foyle, D. C., Ahumada, A. J., Larimer, J., & Sweet, B. T. (1992). Enhanced/synthetic vision systems: Human factors research and implications for future systems. *SAE Transactions: Journal of Aerospace, 101*, 1734-1741.

19. Foyle, D. C., McCann, R. S., & Shelden, S. (1995, in press). Attentional issues with superimposed symbology: Formats for scene-linked displays. In R. S. Jensen (Ed.), *Proceedings of the Eighth International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

20. Denton, G. G. (1980). The influence of visual pattern on perceived speed. *Perception, 9*, 393-402.

21. Johnson, W. W., & Awe, C. A. (1994). The selective use of functional optical variables in the control of forward speed. NASA Technical Memorandum 108849. Moffett Field, CA: NASA.

La Conscience de la Situation en Aéronautique de Combat

Grau J.Y., Menu J.P., Amalberti R.
IMASSA-CERMA
B.P. 73, 91223 Brétigny-sur-Orge Cedex
France

1. SUMMARY
Situation awareness is a key psychological concept improving safety and performance in aircrafts. However, for a long time, any definition was commonly accepted. Now, works in cognitive psychology and ergonomics allow to propose new ways to describe features of situation awareness. Time pressure, risk taking, level of understanding, representation adjustment and anticipation are the elements in order to build dynamical models of situation awareness. These news aspects of situation awareness allow to identify different ergonomic recommendations to drive future interface design and to enhance situation awareness in combat aircrafts.

2. INTRODUCTION
Le concept de conscience de la situation est depuis vingt cinq ans un thème central des études en psychologie de l'aéronautique. Cette évolution est liée à l'importance de la conscience de la situation pour la sécurité et la performance, mais aussi à la complexité croissante des environnements aéronautiques. Que ce soit en aéronautique civile ou militaire, le pilote doit faire évoluer rapidement son aéronef à partir d'informations de plus en plus nombreuses. Il est intéressant de noter que le concept de conscience de la situation était dans un premier temps utilisé au sein de la population des pilotes, avant de l'être par la communauté scientifique. La raison en est que l'ensemble des notions associées à la conscience de la situation n'avait pas d'équivalent dans les thèmes d'étude des différentes disciplines traitant des facteurs humains en aéronautique.
Les premièrs travaux dans la mouvance de la médecine aéronautique n'ont abordé la conscience de la situation que dans le cadre de l'intégration multi-sensorielle et de la désorientation spatiale. Ces approches, trop restrictives, ont rapidement montré leurs limites dans la compréhension des mécanismes sous-jacents à la conscience de la situation.
Au début des années 80, le développement de la psychologie cognitive a permis d'aborder la conscience de la situation à travers la théorie du traitement de l'information et les limitations des capacités intellectuelles des opérateurs. Associant à la notion de charge de travail, les concepts de mémoire de travail et d'attention, de nombreux auteurs comme Fracker (1989) ou Endsley (1989) ont proposé des modèles de la conscience de la situation sous-tendus par la problématique de sa mesure.
Les années 90 ont vu dans le cadre du courant ergonomique, le développement d'une approche complémentaire fondée sur l'intégration des connaissances de la psychologie cognitive et des résultats de l'analyse de l'activité des pilotes s'appuyant sur les études de terrain. A la fois dans le domaine aéronautique civil avec Sarter et Woods (1991) et militaire avec Amalberti et Deblon (1992), le concept de conscience de la situation a été élargi aux caractéristiques dynamiques de la situation de travail. La dimension temporelle de l'activité a été identifiée comme un élément déterminant dans la description des mécanismes étant à l'origine de la prise de conscience de la situation. Les modèles dynamiques de la conscience de la situation qui sont actuellement proposés ouvrent de nouvelles voies à la recherche pour l'optimisation des interfaces homme-système, les systèmes d'aide à l'opérateur et la formation des pilotes.

Dans cet article, après une présentation des définitions et des caractéristiques de la conscience de la situation, nous en proposerons un modèle de gestion dynamique avant d'aborder dans une troisième partie les aides qui pourraient être apportées aux pilotes pour améliorer la conscience de la situation.

3. LA CONSCIENCE DE LA SITUATION

3.1 Définir la conscience de la situation
L'importance de la conscience de la situation n'est plus à démontrer pour un pilote de combat. Mais que sous-entend au juste ce concept et quelles en sont ses caractéristiques? De nombreuses définitions de la conscience de la situation sont proposées dans la littérature, mais permettent elles de la définir avec précision et sans ambiguïté? Si on se réfère à l'article de Sarter et Woods en 1991, définir la conscience de la situation n'est pas aussi aisé qu'on peut le penser a priori.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

Tout d'abord, il semble important de distinguer deux notions qui sont souvent employées l'une pour l'autre : l'orientation spatiale et la conscience de la situation. Menu et Amalberti (1989) définissent l'orientation spatiale comme la capacité à se positionner par rapport à un référentiel fixe que sont les directions verticale et horizontale de l'espace. La conscience de la situation est la capacité à se positionner par rapport à un référentiel relatif constitué par les propriétés dynamiques des objets de l'environnement géographique et tactique. L'orientation spatiale est un mécanisme sous-jacent à la conscience de la situation puisqu'elle la facilite, alors qu'en retour la conscience de la situation ne facilite pas l'appréciation de la situation spatiale.

Les premières définitions de la conscience de la situation que l'on pourrait qualifiées "d'intuitives" sont directement inspirées du langage des pilotes. Elles cherchent à définir les composants de la conscience de la situation et sont plus une liste des informations que le pilote doit connaître ou des buts qu'il doit satisfaire pour réaliser la mission (Mc Kinnon, 1986 et Harwood et al., 1988, cités in Fracker, 1988). La faiblesse de cette approche réside dans le fait qu'on peut avoir des définitions différentes suivant la nature de la mission.
En proposant des définitions basées sur les mécanismes psychologiques, Fracker (1988) et Endsley (1989) ont ouvert la voie à une approche théorique de la conscience de la situation. Ces définitions introduisent les notions d'attention, de niveau d'analyse, de limitation temporelle et de projection dans le futur.
Depuis le début des années 90, les définitions s'orientent de plus en plus sur la dimension dynamique de l'activité des opérateurs dans le cadre de la conduite des situations de contrôle de processus complexe. La définition que propose Amalberti (1995) s'inscrit dans ce courant. Cet auteur définit la conscience de la situation comme la capacité pour le pilote à se former un modèle mental de la situation qui lui permette une action efficace sur cette situation à court, moyen et long terme.

3.2 Caractéristiques de la conscience de la situation mises en évidence par les études de terrain

De nombreux travaux, tant dans le domaine de l'aéronautique que celui des processus industriels, permettent de mieux comprendre certaines caractéristiques de la conscience de la situation.

La première caractéristique est que la conscience de la situation est une construction individuelle, personnalisée. Au delà du paradigme expérimental basé sur la comparaison des pilotes "connus" pour avoir une facilité à construire une bonne conscience de la situation (par exemple ceux qui obtiennent le plus de succès en combat) à ceux qui sont plus souvent en difficulté, il est facile de montrer que pour une même situation, la conscience de la situation est différente entre pilotes. L'analyse de l'activité de huit pilotes ayant des qualifications différentes et effectuant une même mission de pénétration basse altitude en simulateur (Grau et al., 1990), montre qu'aucun pilote n'effectue la mission de la même façon. Les choix tactiques et stratégiques des pilotes sont guidés par leur perception et leur compréhension de la situation qui diffèrent soit au niveau de l'appréciation objective des éléments présents dans l'environnement, soit au niveau des hypothèses d'évolution de la situation dans le court, moyen ou long terme. Cette diversité entre pilotes pose le problème du niveau de finesse avec lequel il faut analyser les mécanismes de la conscience de la situation.

Une deuxième caractéristique tourne autour de la pertinence de la notion de bonne ou mauvaise conscience de la situation. Cette distinction est intuitive mais quelle réalité a-t-elle et comment l'opérationnaliser ? Pendant longtemps, la qualité de la conscience de la situation a été associée à la performance. Mais si on reprend les résultats de l'étude de Grau et al. (1990), on s'aperçoit que malgré des consciences de la situation différentes entre pilotes, les huit pilotes ont rempli correctement la mission. Cela signifierait qu'il y aurait plusieurs bonnes consciences de la situation. La distinction échec / réussite n'est donc pas suffisamment discriminante pour être un indice de la validité objective de la conscience de la situation. L'utilisation d'échelles de discrimination plus fines de la performance est une voie séduisante pour appréhender les écarts entre différentes consciences de la situation mais pose le problème de sa validité opérative vis à vis des pilotes.
Une autre façon d'apprécier la qualité de la conscience de la situation est d'utiliser le paradigme de l'expert. La comparaison pilote / expert permet de pointer facilement les différences et de définir un modèle "idéal" vers lequel doit tendre le pilote novice. Cependant si on observe des pilotes novices, on s'aperçoit que même s'ils connaissent le modèle de l'expert, ce n'est pas pour cela qu'ils l'appliquent car ils n'ont pas les savoir-faire suffisants. En effet, à chaque étape de sa formation, le pilote développe un niveau d'expertise constitué de savoir-faire homogènes et cohérents. La conscience de la situation résulte de ces savoir-faire. Il devient alors délicat de vouloir expliquer la conscience de la situation d'un pilote novice à partir de celle d'un pilote expert. La qualification de la conscience de la situation doit se faire en fonction de chaque pilote. Cette démarche est proche de celle que Reason (1990) préconise dans un autre domaine du comportement humain quand il définit l'erreur humaine comme un écart à l'intention de l'opérateur. L'appréciation de la validité de la conscience de la situation ne se fait plus par rapport à une référence externe mais par rapport à une référence interne propre à chaque opérateur.

La troisième et dernière caractéristique est constituée par les liens qui existent entre conscience de la situation et contraintes temporelles. Dans les situations dynamiques, le temps est au cœur de l'organisation de l'activité. Les travaux sur la gestion du temps dans les activités complexes (de Keyser, 1991 ; Grau, 1993) décrivent différentes stratégies d'adaptation de l'activité aux contraintes temporelles. Une de ces stratégies est la possession par l'opérateur de plusieurs registres de fonctionnement pour une même tâche. Chaque registre de fonctionnement est associé à un niveau de pression temporelle. Sperandio (1984) a décrit avec précision des registres similaires dans le contrôle aérien lorsque la charge de travail croît.
Une étude menée récemment dans notre laboratoire s'est intéressée à la conscience de la situation lors du combat air-air. Onze pilotes ont effectués les mêmes scénarios de combat en simulation. Dans un premier temps, chaque pilote après le vol a décrit sa perception de la situation. Dans un second temps, les informations présentées au pilote pendant le vol lui ont été remontrées en salle de "debriefing" avec tout le temps nécessaire pour les étudier. Dans un troisième temps, chaque pilote a décrit la conscience de la situation qu'il avait à l'issue de cette nouvelle présentation et comment il l'évaluait par rapport à celle qu'il avait construite au cours du vol. L'évaluation était une évaluation subjective sur une échelle à 5 niveaux allant du niveau 1 où la conscience de la situation pendant le vol est totalement erronée jusqu'au niveau 5 où elle n'a pas évolué à l'issue du rejeu. Les résultats de cette auto-évaluation montrent d'une part que la conscience de la situation pendant le vol est dans tous les cas moins exhaustive que lors du rejeu ; et d'autre part que la conscience de la situation est évaluée aux niveaux 1, 2 et 3 lorsque la pression temporelle est élevée. A travers ces résultats, il est possible d'objectiver pour chaque pilote, différents niveaux de conscience de la situation dont la mise en œuvre dépend de la pression temporelle.

4. CONSCIENCE DE LA SITUATION ET COMPRÉHENSION

La compréhension est un thème majeur des études en psychologie cognitive depuis de nombreuses années. Les modèles développés à partir de données expérimentales de laboratoire, fournissent une base explicative aux mécanismes de la conscience de la situation. Cependant, laboratoire et environnement de travail sont des mondes où les activités des sujets répondent à des contraintes bien différentes. La connaissance des résultats expérimentaux est essentielle tout en sachant qu'elle ne permet pas de tout expliquer.

Avoir conscience de la situation, c'est la comprendre, c'est à dire en construire une représentation qui s'inscrit dans l'activité du pilote. Cette représentation est fonctionnelle (Richard, 1985 ; Leplat, 1988 ; Ochanine, 1981) dans le sens où elle est finalisée, transitoire, sélective, laconique et évolutive. La construction de cette représentation permet de donner de la cohérence aux faits qui constituent la situation de travail, mais aussi de satisfaire l'objectif que se fixe l'opérateur pour atteindre un but. Deux mécanismes prévalent dans la construction de la représentation.

Le plus connu est celui de la particularisation d'un schéma. Le concept de schéma proposé par Norman (1983) vise à décrire les connaissances sous forme de blocs de connaissances, indépendants des autres connaissances, disponibles tels quels en mémoire et comprenant à la fois des concepts, des raisonnements, des actions et les relations entre ces différents éléments. Le schéma présente un degré de généralité et son utilisation nécessite de le particulariser, c'est à dire d'en instancier certaines variables. Le "décollage" est un bon exemple de schéma : le pilote dispose en mémoire des séquences d'actions, des notions et des opérations à effectuer depuis le lâcher des freins jusqu'à la rejointe d'un niveau de vol. Cependant pour appliquer ce schéma dans un contexte donné, le pilote doit insérer certaines valeurs qui vont permettre de l'opérationnaliser (masse et configuration de l'aéronef, direction et vitesse du vent, température extérieure, type de décollage, traffic aéronautique, etc ...).
La sélection du schéma peut se faire par un mécanisme "top-down" à partir du nom du schéma. Le schéma ainsi sélectionné va créer des attentes qui vont guider les prises d'information du pilote et ses raisonnements. Au décollage, le pilote sait qu'il va devoir vérifier sa vitesse et son accélération à certains moments du roulage, mais il connaît aussi les valeurs attendues. La prise d'information est alors un mécanisme actif, attentionnel, dirigé par le but. Des informations pertinentes, non conformes aux attentes du schéma peuvent passer inaperçues et être négligées car hors du champ de l'attention du pilote.
Un autre mode de sélection du schéma est un mécanisme 'bottom-up" à partir des données de la situation. Les informations prélevées dans l'environnement sont structurées, organisées par le pilote jusqu'à ce qu'elles puissent être comparées à un schéma stocké en mémoire. La situation est alors comprise comme étant celle du schéma. Ce mécanisme est fréquent face à un événement non anticipé. Il permet de ramener la situation à quelque chose de connu et de faire un diagnostic. La difficulté de la reconnaissance du schéma tient à la quantité et au saillant des informations disponibles dans l'environnement. Une fois le schéma sélectionné, il permettra d'orienter l'activité du pilote.
Pour faire face à la complexité des situations rencontrées (flux informationnel important, pression temporelle, imprécision de certains buts, incertitude sur les informations et interruptions fréquentes), le pilote ne peut utiliser de façon exhaustive les mécanismes de sélection décrits ci-dessus. Il va développer des

raccourcis qui vont lui permettre de traiter plus rapidement les informations. Reason (1988) décrit deux types de raccourcis : le "pattern matching" qui consiste à faire une comparaison sur un nombre limité d'éléments du schéma, et le "frequency gambling" qui consiste à faire la sélection parmi les schémas les plus fréquents car les plus disponibles en mémoire. Dans les deux cas, il va y avoir gain de temps mais risque de sélection trop rapide d'un schéma qui va conduire à une mauvaise représentation de la situation, Reason parle de "rationalité limitée".

L'intérêt du schéma se situe dans ses qualités diagnostiques, mais aussi dans les propriétés d'anticipation qu'il procure à l'opérateur en permettant de faire des inférences sur le futur. Une étude visant à expliquer les différences d'activité entre pilotes dans une même situation (Grau et al., 1992) a mis en évidence la place des schémas dans l'élaboration de la conscience de la situation. Huit pilotes ont été confrontés à un même scénario papier d'engagement air-air. A partir des mêmes informations, chaque pilote a construit une structure de connaissances qui lui permet de porter un diagnostic sur la situation mais aussi de faire des inférences sur son évolution. Les diagnostics étaient peu différents entre pilotes, ce qui était vraisemblable en raison des conditions de l'expérimentation. Par contre, les anticipations à court, moyen et long terme de l'évolution de la situation qui engageaient directement les actions des pilotes présentaient des différences considérables. Les plans d'interception proposés par les pilotes pouvaient être radicalement opposés. Cette étude permet de préciser deux points : la sélection des schémas est plus complexe que la seule sélection des informations et l'importance des schémas dans l'élaboration de la conscience de la situation.

Un deuxième mécanisme de représentation est la construction d'une structure particularisée de la situation. Ce mécanisme est mis en œuvre lorsque l'opérateur ne possède pas de schéma adapté à la situation. On se retrouve dans une situation dite de résolution de problème, pour laquelle comprendre c'est construire un espace de recherche du problème dans lequel le pilote va trouver un cheminement entre l'état initial et l'état but (Newell et Simon, 1972). Ce mécanisme coûteux, nécessite des ressources mentales importantes. En conséquence, il est rarement mis en œuvre dans les situations à forte pression temporelle car le pilote préfère utiliser le mécanisme précédent.

Ces mécanismes fondamentaux de la compréhension sont la base de la prise de conscience de la situation. Cependant, ils ne peuvent être appliqués tels quels à une situation de contrôle de processus continu et rapide comme le pilotage d'avions d'armes. Les études ergonomiques fondées sur l'analyse de l'activité de l'opérateur apportent un point de vue complémentaire sur le concept de conscience de la situation

5. VERS UN MODÈLE DYNAMIQUE DE LA CONSCIENCE DE LA SITUATION

5.1 Le niveau de compréhension

Dès 1988 dans sa définition de la conscience de la situation, Fracker introduit le concept de niveau d'abstraction de la compréhension pour expliquer la conscience de la situation. Pour cet auteur, les hauts niveaux d'abstraction sont les buts de la mission, et les niveaux les plus bas, les variables de la situation à un moment donné. Plus le pilote comprend à haut niveau, meilleure est la compréhension des bas niveaux.

Une autre façon d'envisager le niveau de compréhension est de le définir comme le degré de particularisation de la représentation. Schank dans le cadre de la théorie des schémas a décrit des organisations hiérarchiques et emboîtées de structures, les schémas, allant des niveaux les plus spécifiques aux niveaux les plus abstraits. L'objet de ces représentations est le même mais avec un degré de raffinement ou de complétude plus ou moins grand. Pour Rasmussen (1990), le raffinement peut se faire suivant deux dimensions : une dimension abstrait/concret qui concerne les concepts utilisés et une dimension tout/parties qui concerne le niveau de détail choisi. Suivant la première dimension, les représentations peuvent varier pour un même objet puisque la représentation s'inscrira dans des logiques et des raisonnements différents. Elles dépendront des connaissances de l'opérateur et de sa capacité à les utiliser dans une situation donnée. La seconde dimension est quand à elle beaucoup plus en rapport avec la possibilité pour l'opérateur de traiter une somme de données plus ou moins grande.

Aux niveaux les plus spécifiques, la représentation résulte du traitement de l'ensemble des données présentes dans l'environnement. Elle permet d'avoir des raisonnements précis directement applicables à un niveau tactique. Aux niveaux les plus abstraits, la représentation permet une identification générique de la situation à partir de laquelle les pilotes ne pourront effectuer que des traitements stratégiques. A chaque situation, il correspond un niveau optimal de compréhension qui résulte d'une utilisation optimale de ces caractéristiques. Suivant les phases de mission, les besoins de compréhension des pilotes nécessiteront différents niveaux de compréhension. Le degré de particularisation de la représentation permet de répondre aux exigences cognitives de la tâche en favorisant la construction de l'activité.

Le passage d'un niveau de compréhension à un autre résulte soit d'un mécanisme d'abstraction dans lequel la représentation est évoquée de façon globale, soit d'un mécanisme de raffinement où la décomposition en buts et sous-buts qui permet au pilote de se représenter la situation à différents niveaux de détails. Dans une tâche, le niveau de compréhension de la situation est fonction des savoir-faire (Amalberti et Valot, 1990) : les pilotes analysent à un niveau fin les parties les plus difficiles à

comprendre alors qu'ils laissent les parties bien connues à un niveau très général.

5.2 Être conscient de la situation, c'est ajuster sa représentation de la situation

Si la compréhension de la situation peut-être définie comme la construction d'une représentation, on peut aussi l'envisager dans les situations de contrôle de processus continu comme un mécanisme d'ajustement permanent de la représentation. Le processus de compréhension n'est plus guidé par la survenue de situations particulières ou non anticipées. Il correspond à une mise à jour constante, dépendante de l'historique de la situation et des buts à atteindre. Cette idée est reprise en d'autres termes par Sarter et Woods (1991) dans l'aviation civile qui distingue évaluation de la situation et conscience de la situation. La conscience de la situation correspond à l'intégration de connaissances provenant d'évaluations périodiques de la situation. L'évaluation de la situation est quand à elle décrite comme un processus complexe de perception et de comparaison limité par les capacités de la mémoire de travail et des ressources attentionnelles (Endsley, 1989). Sous cette acception, l'évaluation de la situation se rapproche de la définition de la compréhension qui a été envisagée au paragraphe 4. La conscience de la situation peut alors être décrite comme un processus indépendant de la compréhension qui vise à gérer en permanence les écarts au réel. On ne définit plus la conscience de la situation comme la construction d'une représentation mais comme un ajustement, une mise à jour de la représentation en fonction de l'évolution de la situation courante (Wickens, 1992). Comprendre consiste à intégrer les nouvelles informations à la représentation en cours. Cette intégration peut se faire par assimilation (enrichissement et évolution de la représentation en cours) ou accommodation (modification de la représentation en cours). Le second mode d'intégration est peu rencontré en pilotage car coûteux. En fait, le pilote préfère ramener la situation à quelque chose de connu. L'ajustement obéit à un principe d'économie (on ne cherche pas à avoir la représentation la plus exhaustive mais la plus fonctionnelle) et de persévérance (tendance à favoriser la représentation initiale).

A travers le mécanisme d'ajustement, on voit les liens étroits qui existent entre les différents processus cognitifs du traitement de l'information. Dans des études sur le diagnostic et la prise de décision, Hoc et Amalberti (1994) ont montré que les opérateurs favorisent les décisions ou les informations pour lesquels ils avaient des possibilités d'action. Il existe un lien très fort entre répertoire de réponse, mise en situation et type de diagnostic évoqué. La conscience de la situation est ainsi régulé par un ajustement permanent qui s'appuie sur la logique de l'action. L'imbrication de la conscience de la situation avec la décision et l'action amène à réfléchir sur l'ordonnancement dynamique de ces processus. Longtemps considérés comme des phénomènes séquentiels, les raccourcis et les aller-retour qui existent entre ces processus laissent plutôt penser à un fonctionnement en "parallèle" beaucoup plus adapté aux contraintes dynamiques des situations de travail.

5.3 La Gestion de compromis à la base de la conscience de la situation

Lorsqu'on étudie la conscience de la situation d'un pilote, on constate que cette dernière est constituée d'un ensemble d'éléments que l'on peut facilement classer sous différentes rubriques. Par exemple, les objets suivants peuvent être identifiés dans une mission de défense aérienne : le cadre tactique air-air, le cadre tactique sol-air, la gestion des capteurs de l'aéronef, la gestion de l'armement, le pilotage de l'aéronef et la gestion de la mission. La description de ces différents objets conduit à se poser une question importante pour l'acquisition de la conscience de la situation : doit-on parler de la conscience de la situation comme une représentation supra-ordonnée d'un ensemble de représentations de niveaux inférieurs ou comme une construction globale et unique de l'ensemble des données présentes dans l'environnement ? Les données obtenues à partir de la formalisation des connaissances manipulées au cours d'une mission de pénétration en basse altitude (Grau et al, 1990) mettent en avant un fonctionnement mental basé sur des représentations permanentes portant sur des objets différents qui évoluent en parallèle. La conduite d'un aéronef exige que le pilote gère un grand nombre d'informations et de connaissances. En raison de la limitation des capacités humaines, le pilote ne peut à tout moment avoir une représentation intégrant l'ensemble des informations présentes. Il doit adopter un mode de traitement qui soit un compromis coût/performance. La mission est découpée en sous-ensembles ou objets pour lesquels le pilote a une représentation. Chaque représentation permet de faire des anticipations. Il est ainsi possible au pilote de répartir ses ressources cognitives en fonction des anticipations, pour les porter sur les objets les plus pertinents pour la tâche. Il dégage pour chaque représentation des fenêtres temporelles dans lesquelles il peut effectuer des prises et traitements d'informations pour effectuer les ajustements nécessaires. Tous les objets ont une représentation mais à un niveau de compréhension qui varie en fonction des buts du pilote. Ainsi, par exemple le niveau de conscience adopté par le pilote pour le bilan pétrole sera différent avant de passer les lignes ennemis (quantité de l'écart par rapport à ce qui est planifié), lors du passage des lignes (il y a assez de pétrole ou non), en territoire ennemi (alarme de consommation du pétrole minimum programmé en préparation), lors d'un engagement (alarme que la réserve de combat est épuisée), au cours du retour terrain (bilan pétrole avec équivalence en temps de vol, niveau de vol et vitesse à prendre) et lors d'un

déroutement (bilan très fin avec consommation des réserves de sécurité). Une conscience de la situation pertinente traduit les capacités du pilote a assigné aux différentes représentations en cours le niveau de compréhension optimal. L'acquisition et le maintien de la conscience de la situation peuvent être envisagés comme un mécanisme de gestion des niveaux de compréhension des différents objets de la mission.

La détermination des niveaux de compréhension des différents objets a deux origines. La première est celle des objectifs de la tâche. Il est facile de comprendre que suivant les phases de la mission, les exigences de la tâche pour les différents objets évoluent. La deuxième origine est liée aux connaissances et savoir-faire du pilote. Tout pilote, avec l'acquisition de l'expérience, développe des connaissances sur ses propres connaissances et capacités. Il connaît ses points forts, ses points faibles. Ce type de connaissances est appelé métaconnaissances, et permet au pilote de contrôler le niveau de compréhension utile. Il permet d'adapter son activité et ses comportements aux exigences de la tâche. Ainsi, un pilote peut faire des choix en fonction de ce qu'il se sait capable de faire dans le temps qui lui est imparti et non pas seulement, en fonction de ce qu'il se sait capable de faire au mieux. Connaître la réponse à une situation est utile, mais savoir si on peut l'appliquer ou non est la question clé à se poser. L'élaboration des métaconnaissances est complexe et résulte de la mise en situation des connaissances du pilote avec la pression temporelle et la prise de risque. La notion de risque est complexe. Wilde (1988) estime que la prise de risque liée au choix d'un comportement est subordonnée à la confrontation de deux représentations : le risque préférentiel et le risque perçu. Le risque préférentiel est le niveau subjectif de risque où le sujet estime que le rapport entre les bénéfices escomptés et les coûts liés au comportement est maximal ; c'est le risque attendu. On peut aussi le définir comme le niveau que le pilote estime pouvoir et devoir prendre pour atteindre le but qu'il s'est fixé. Le risque perçu est la probabilité subjective d'occurrence d'une dégradation et l'évaluation de sa gravité. En aéronautique, Amalberti (1992) a développé 2 notions proches de celles de Wilde : le risque interne qui est la probabilité de dépasser ses ressources cognitives, et le risque externe qui est la conséquence d'un comportement sur la performance. Dans la tâche de pilotage, le pilote cherche en permanence à contrôler le risque interne, parfois au détriment du risque externe qu'il maîtrise mieux, du moins subjectivement. Le risque devient alors un guide qui structure le cours de l'activité et les niveaux de compréhension des différents objets.

Dans la conscience de la situation, les métaconnaissances interviennent directement dans la gestion du compromis entre niveau de compréhension et anticipation. Plus le niveau de compréhension est élevé, meilleure sera l'anticipation et plus le pilote pourra gérer au mieux la répartition de ses ressources cognitives entre les différentes représentations. Cette situation qui peut être qualifiée "d'idéale" est rarement possible en aéronautique de combat car le pilote ne dispose pas du temps suffisant pour prendre et traiter l'information nécessaire. A l'inverse, un niveau de compréhension faible est responsable d'une faible capacité d'anticipation qui peut conduire à des situations dangereuses. Dans la pratique, le pilote est le plus souvent dans une situation intermédiaire où la conscience de la situation résulte de la détermination des différents niveaux de compréhension des objets de la mission en fonction des exigences de la tâche et de ses propres capacités.
Lorsque les exigences de la tâche amènent le pilote à monopoliser ses capacités sur un seul objet, cela se fait au détriment des autres objets. Ainsi, Amalberti (1995) décrit comme un mécanisme essentiel de la conscience de la situation, l'acception par le pilote de "ne pas comprendre" car les ressources cognitives qu'il devrait allouer à la compréhension sont incompatibles avec la poursuite de la tâche. Ce mécanisme a été mis en évidence à travers deux types de comportement :

- le premier consiste à différer la compréhension et la réponse d'un problème plutôt que de risquer de dépasser ses ressources attentionnelles.
- le second montre que lorsque les exigences croissent, le pilote peut être amener à redéfinir sa tâche pour la rendre compatible avec ses capacités. Soit il la redéfinit en restant dans des limites qu'il juge acceptable, c'est par exemple une moins bonne précision de l'heure de passage sur une cible ; soit le pilote va redéfinir les objectifs de la mission en délaissant certains buts pour ne garder que ceux qui lui semblent pertinents. Le pilote va en quelque sorte négliger toute une partie de sa tâche (surveillance du ciel par exemple) pour ne se consacrer qu'à celle qu'il se sent capable de réaliser (navigation). La performance est dégradée par rapport à la tâche prescrite, mais le pilote fait bien ce qu'il a choisi de faire.

La conscience de la situation est un mécanisme complexe qui résulte de nombreux compromis cognitifs. Elle est le résultat de la construction de plusieurs représentations qui s'élaborent simultanément et en permanence. Le niveau de compréhension associé à chaque représentation est la clé cognitive de la conscience de la situation. Le pilote ne peut avoir un niveau de compréhension "maximal" pour chaque représentation. Il est obligé de choisir un "profil" des niveaux de compréhension des différentes représentations propre à la situation rencontrée. Au cours des différentes phases de la mission, le pilote manipule les niveaux de représentation en changeant de niveau d'abstraction. Le garant d'une conscience de la situation réside non pas dans le traitement du plus grand nombre d'informations, car on a vite fait de dépasser les capacités cognitives, mais bien dans le réglage cognitif

des niveaux de compréhension des différentes représentations des objets de la mission. Ce réglage cognitif est une habilité qui s'apprend avec l'expérience et qui est l'apanage du pilote expert. Elle est cependant remise en question à chaque vol et nécessite une attention constante pour garantir la sécurité et la performance.
D'un point de vue ergonomique, la mise à plat des mécanismes participant à l'élaboration de la conscience de la situation permet d'envisager de nouvelles recommandations sur les aides qui pourraient être apportées à l'opérateur pour améliorer ses connaissances et son évaluation de sa représentation de la situation.

6. AMÉLIORER LA CONSCIENCE DE LA SITUATION

Les pistes pour améliorer la conscience de la situation sont multiples. Elles comprennent l'ensemble des actions que l'on peut mener dans le domaine des facteurs humains et de l'interface homme-machine. L'assistance au pilotage et les problèmes liés à la formation seront abordés dans cet exposé.

6.1 L'assistance au pilotage

Un premier concept d'assistance est de présenter les informations conformément à leur représentation dans la conscience de la situation. Ce concept développé par Zachary (1989) sous le nom d'aide à la représentation vise à présenter l'information sous une forme où elle soit directement compatible avec les représentations mentales. Ainsi, une représentation spatiale en 3 dimensions garantit une meilleure compatibilité pour les tâches visant à définir des trajectoires où à positionner des éléments extérieurs, qu'une présentation verbale sous forme de conseils. De même, lors de la présentation d'informations qui résultent de calculs effectués par les systèmes, les données présentées doivent être compatibles avec la logique d'utilisation du pilote et non avec la logique physique du système.

Un second concept est de présenter au pilote toute et seulement l'information pertinente. C'est un pari ambitieux que de réaliser un tel système. De nombreux travaux vont dans ce sens, mais l'échec partiel des tentatives réside dans la difficulté à appréhender la globalité des connaissances d'un pilote. En effet, un grand nombre de ces connaissances est difficilement accessible à la conscience soit parce qu'elles sont automatisées (Schiffrin et Schneider, 1977) soit parce que ceux sont des connaissances tacites qui relèvent d'apprentissages implicites (Berry et Broadbent, 1984). Ces résultats ont conduit à faire évoluer le concept pour configurer l'interface en fonction des intentions du pilote. Développé dans la cadre du programme français "Copilote Électronique", ce concept vise à identifier à partir des actions pilote, de l'évolution des paramètres de l'aéronef et de la connaissance de l'environnement, les intentions du pilote en utilisant les techniques de l'intelligence artificielle. Après avoir extrait l'expertise pilote par entretiens et observations, il est possible de structurer cette connaissance sous forme de schémas qui permettent d'inférer les intentions probables du pilote. L'avantage de cette approche est de déduire à partir des intentions et des données sur la situation, la charge de travail du pilote et de ne proposer que les informations qu'il pourra traiter en fonction des niveaux de compréhension de chaque objet adéquat. Par exemple, la modulation des conseils en fonction du temps disponible ou des dialogues à plusieurs vitesses dans les situations d'urgence sont autant de choix dans un système qui facilitent la compréhension par le pilote car conformes aux mécanismes cognitifs.

Une autre recommandation concerne l'introduction dans les aéronefs de systèmes d'aide au pilotage. Développée dans les programmes "pilot's associate" aux USA ou de "Copilote Électronique" en France, cette recommandation a pour but de doter le pilote d'un système qui le prévienne de situations dangereuses (perte de la suprématie aérienne) et qui le conseille sur les choix tactiques (choix de la manœuvre de combat la plus appropriée en fonction de la situation tactique et des paramètres de l'aéronef). L'intérêt est de sensibiliser le pilote sur des aspects de la situation qu'il a délaissé ou qu'il a mal interprété. La réalisation de tels systèmes doit prendre en compte dès les phases de conception, les contraintes liés aux spécificités de l'élaboration de la représentation mentale. Une des principales contraintes est d'avoir un fonctionnement proche de celui du pilote. On parle de comportement "human like" qui permet de rendre le système "transparent". Le pilote comprend ainsi les opérations effectuées par le système et développe un sentiment de confiance, évitant des interprétations magiques (Amalberti, 1992). La réalisation d'un tel système nécessite de disposer d'un bon modèle de l'activité du pilote et d'une expertise homogène afin d'éviter des comportements incohérents. Le pilote doit avoir une représentation du type d'aides que peut lui apporter le système afin de ne pas être surpris par ce dernier. Cela ne serait pas le cas si l'expertise introduite dans le système était une mosaïque de connaissances extraites chez différents experts. D'un point de vue cognitif, le bénéfice d'aide à la représentation est de favoriser les anticipations, un système réactif n'apporterait que peu d'aide. Pour cela, le système doit lui-même être capable d'anticipation afin qu'il puisse dialoguer en synergie avec le pilote, et accroître la conscience de la situation.

Un dernier type d'aide est celui qui peut être apporté par les systèmes qui facilitent dès la préparation de mission l'anticipation de la charge de travail et des niveaux de compréhension au cours de la mission.

6.2 L'entraînement

L'entraînement est un élément incontournable dans l'acquisition des connaissances qui facilitent l'élaboration de la conscience de la situation. Les évolutions théoriques sur l'acquisition des capacités cognitives montrent que si pendant longtemps, on a pensé qu'il serait possible d'apprendre des connaissances génériques sur comment prendre une décision ou comment construire une représentation (Hunt et Rouse, 1981), on préfère actuellement aider le pilote à acquérir un catalogue d'exemples auquel il puisse faire référence. Il est donc nécessaire que le pilote puisse explorer toutes les ressources de son aéronef et qu'il les confronte à des situations tactiques de plus en plus complexes afin d'acquérir les schémas de raisonnements et d'actions, mais aussi les métaconnaissances qui faciliteront la gestion des compromis cognitifs indispensables à l'élaboration de la conscience de la situation.

Un dernier aspect de la conscience de la situation, qui n'a pas été abordé jusqu'à maintenant, est celui de la prise de conscience de la situation au niveau collectif lorsque plusieurs pilotes sont engagés dans des tâches différentes pour atteindre un même objectif. Si ces situations sont l'apanage de l'aviation de transport, il n'en est pas moins vrai qu'on les rencontre fréquemment en aviation de combat au sein d'un équipage d'un avion biplace, mais aussi au sein d'un dispositif. La qualité de la conscience collective de la situation nécessite de développer chez les pilotes des attitudes de communication. Des programmes de formation à ces attitudes existent et sont regroupés sous les termes génériques de Cockpit Resource Management (CRM). Mis au point sous l'impulsion des travaux de Wiener (1989) sur les conséquences des mauvaise communications dans les glass-cockpit, ils sont maintenant généralisés dans toutes les grandes compagnies aériennes et des programmes spécifiques sont développés pour l'aviation de combat.

7. CONCLUSION

Les travaux sur la conscience de la situation sont d'un intérêt essentiel pour la sécurité et la performance en aéronautique militaire. Le concept de conscience de la situation a considérablement évolué au cours des ces dernières années grâce à l'apport de la psychologie cognitive et des résultats des études de terrain. La conscience de la situation n'est plus envisagée comme une propriété de la situation. Elle dépend principalement de ce que le pilote veut faire dans la tâche et du niveau de risque qu'il désire prendre. L'accident ou l'échec de la mission qui traduisent une désadaptation profonde de la conscience de la situation ne doivent plus être considérés comme une absence de représentation ou une représentation erronée de la situation, mais plutôt comme un dérapage des mécanismes d'auto-évaluation de ce qui peut être fait sans trop de risques avec la représentation disponible. Les caractéristiques des modèles dynamiques de la conscience de la situation permettent de faire des recommandations pour améliorer la conscience de la situation des pilotes dans les cockpits.

BIBLIOGRAPHIE

Amalberti R. (1992). Modèles d'activité en conduite de processus rapides : implications pour l'assistance à la conduite. Thèse doctorat "psychologie des processus cognitifs". Université Paris VIII.

Amalberti R. (1995). La conduite des systèmes à risque. PUF, Paris (à paraître).

Amalberti R. and Deblon F. (1992). Cognitive modelling of fighter aircraft's control process : a step towards intelligent onboard assistance system. International journal of man-machine studies, 36, 639-671.

Amalberti R. and Valot C. (1990). Champs de validité pour une population de pilotes de l'expertise de l'un d'entre eux. Le travail humain, 53, 4.

Berry D.C. and Broadbent D.E. (1984). On the relationship between task performance and associated verbalizable knowledge. Quartely journal of experimental psychology, 36A, 209-231.

Endsley M.R. (1989). A methodoly for the objective measurement of situation awareness. AGARD conference proceedings n° 478. AGARD, Neuilly-sur-seine, France.

Fracker M.L. (1988). A theory of situation assessment : implications for measuring situation awareness, in Proceedings of the human factor society 32nd. annual meeting (pp 102-106). Santa Monica, CA : Human Factor Society.

Fracker M.L. (1991). Measures of situation awareness : Review and future directions. AL Report n° AL-TR-1991-0128. Wright-Patterson AFB, OH : Armstrong Laboratory.

Grau J.Y. (1993). Error in time management during fighting mission, in proceedings of the 7th. international symposium on aviation psychology. Columbus, OH, april 1993.

Grau J.Y. and Amalberti R. (1990). Le transfert de connaissances : enjeux pour l'aéronautique. Revue de médecine aéronautique et spatiale, XXIX, n°113.

Grau J.Y. and Bellet T. (1992). Bases théoriques de l'erreur humaine en situation de pilotage. Rapport CERMA 92-25, Décembre 1992.

Hoc and Amalberti R. (1994). Diagnostic et prise de décision dans les situations dynamiques. Psychologie française, 39-2.

Hunt R. and Rouse W. (1981). Problem solving skills of maintenance trainees in diagnosing faults in simulated powerplants. Human factors, 23(3), 317-328.

Keyser de V. (1991). Temporal reasoning in continuous processes : segmentation and temporal reference systems. Third european conference "Cognitive science approaches to process control". Cardiff, UK, 2-6 september 1991.

Leplat J. (1988)). Task complexity in work situations, in Goodstein, Andersen and Olsen (eds). Tasks, errors and mental models. Taylor and Francis publish. London.

Menu J.P. and Amalberti R. (1989). Les déterminants de l'appréciation de la situation tactique et le développement de systèmes d'aides ergonomiques. AGARD conference proceedings n° 478. AGARD, Neuilly-sur-seine, France.

Newell A. and Simon H. (1972). Human problem solving, Englewood cliffs, New-York, Prentice Hall.

Norman D. (1983). Some observations on mental models, in Mental models, Stevens & Gentner (eds). Hillsdale : L. Erlbaum.

Ochanine D. (1981). L'homme dans les systèmes automatisés. Dunod, Paris.

Sarter N.B. and Woods D.D. (1991). Situation awareness : a critical but ill-defined phenomenon. International journal of aviation psychology, 1(1), 45-57.

Rasmussen J. (1990). Taxonomy for cognitive work analysis, in Rasmussen, Brehmer, de Montmollin and Leplat 'eds). Proceedings of the first MOHAWC project, Vol 1, Riso : Roskilde.

Reason J. (1988). Framework models of human performance and error : a consumer guide, in Tasks, errors and mental models, L. Goodstein and A.H. Olsen (eds). London, Taylor and Francis.

Reason J. (1990). Human error. Cambridge university press.

Richard J.F. (1985). La représentation du problème. Psychologie Française, 30, 3/4, 277-284.

Schank R.C. and Abelson R. (1977). Scripts, plans, goals and understanding. Hillsdale : L. Erlbaum.

Schiffrin R.M. and Schneider W. (1977). Controlled and automatic human information processing : perceptual learning, automatic attending and a general theory. Psychological Review, 84, 155-171.

Spérandio C. (1984). L'ergonomie du travail mental. Masson, Paris.

Wickens C. (1992). Workload ans situation awareness : an analogy of history and implications, INSIGHT : the visual performance technical group newsletter, 14.

Wiener E.L. (1989). Human factors of advanced technology ("glass-cockpit") transport aircraft. NASA contractor report n° 177528. Moffet Field, CA : NASA-Ames Research Center.

Zachary W. (1988). Decisions support systems : designing to extend the cognitive limits, in Handbook of human-computer interaction, Helander M. (eds), Elsevier science : Amsterdam (North-holland).

ATTENTIONAL CONTROL AND SITUATIONAL AWARENESS IN A COMPLEX AIR COMBAT SIMULATION

M.A. Vidulich

Armstrong Laboratory
Human Engineering Division
AL/CFHP
2255 H Street
Wright-Patterson AFB OH 45433-7033
USA

A.L. McCoy
M.S. Crabtree
Logicon Technical Services, Inc.
Dayton, Ohio 45431, USA

SUMMARY

Situational awareness (SA) refers to a pilot's cognitive understanding of the mission situation. SA is complex and difficult to define because it involves a wide variety of cognitive processes. However, the present research hypothesized that one cognitive process, attentional control, would be key to understanding and enhancing pilot SA. To test this hypothesis a training procedure that was expected to improve an individual's attentional control was performed by one experimental group. Another group performed a placebo training procedure. Both groups received their training embedded within a larger program of performing complex air combat simulation missions. Performance and SA of the groups was measured both before and after the attentional control or placebo training. As expected, the group that received the attentional control training showed a greater improvement in performance than did the placebo control group. However, although the SA metrics appeared to be sensitive to an SA manipulation within the simulation, there did not appear to be any SA benefit associated with the attentional control training.

INTRODUCTION

Situational awareness (SA) has become a popular topic within the human factors community. As with many topics, the definition of SA is not universally agreed upon. One popular definition by Endsley (Ref 1) regards SA as (p. 97):

"the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

A United States Air Force Air Staff process action team (Ref 2) defined SA as (p. 6):

"a pilot's (or aircrew's) continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception."

In a review of these and numerous other definitions that have been proposed in the literature, Dominguez (Ref 3) suggested that a consensus definition might be (p. 11):

"Continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events."

From these definitions it is clear that SA is a complex issue that involves a wide variety of cognitive processes to create and maintain. In order to understand SA well enough to accurately predict or assess it, it is necessary to determine the relevance of various cognitive processes to SA in operational contexts of interest.

The present paper presents a study to determine the role of attentional control in performance and SA during a complex air combat simulation.

Attentional Control, Performance, and SA

Recent research has demonstrated that attentional control is important to flying skill in beginning military aviation. Gopher, Weil, and Bareket (Ref 4) demonstrated that the use of an emphasis-change training protocol within an arcade-style video game transferred to flight training in the Israeli Air Force. Hart and Battiste (Ref 5) replicated this finding in a comparable study using U.S. Army helicopter pilot trainees. In the Hart and Battiste study the effect of the emphasis-change training with the arcade-style game was demonstrated to be greater than the effect observed in a placebo control group that performed a video game possessing higher face validity to the helicopter task but lacking the emphasis-change training.

The possibility that 10 hours of experience with a video game (and the emphasis-change training routine) could significantly improve the performance of pilot trainees has potentially important implications for the design of training programs and our understanding of the cognitive processes involved in military flight. In attempting to explain the results of both his own and Hart and Batiste's studies, Gopher (Ref 6) suggested that the experience of shifting attention within the demanding context of the arcade-style game used for training educated the subjects about the potentiality and effectiveness of developing appropriate attentional strategies for complex environments. In other words, the subjects developed a generic attentional control skill that could effectively be transferred to the flight training environment.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

The present paper presents an attempt to replicate and extend the previous results. Two aspects of the present study will be focused on in this paper: One, can the performance enhancing effects of the attentional control training be replicated within a laboratory simulation of complex air combat? Two, assuming that the performance effect is replicated, is the benefit of the attentional control training associated with an improvement in subject SA during the experimental trials? The hypothesis was that if the training improved a generic attentional control skill, then it was expected that the more adaptive attentional control would provide more effective information intake and produce improved situation awareness.

METHOD

Two groups of college-age subjects were used in the experiment. Each group consisted of 16 subjects. Assignment to the two groups was quasi-random, with the constraints that the two groups have approximately the same proportions of male and female and the same proportion of individuals with pilot or video game experience.

Groups

The two groups differed in the nature of the subsidiary training each received during the course of the experiment. One group performed the same arcade-style game used by Gopher et al. (Ref 4) and Hart and Battiste (Ref 5), called *Space Fortress*, and received the emphasis-change training regime used in the previous studies. This group was designated Group SF.

The second group received a placebo training program that was called *IQ Builder*. This group was designated Group IQ.

Experimental Task

All subjects performed the same experimental tasks throughout the experiment. The task was a commercial air combat simulation package called *Chuck Yeager's Air Combat*. The simulation required that subjects fly one aircraft in air-to-air combat with other aircraft. The number and type of other aircraft (friendly or enemy), starting situation, and mission goals varied across trials of the simulation. Each simulation trial started with an on-screen briefing of the starting simulation and mission goals and ended with feedback about performance and mission accomplishment. The simulation package provided three "eras" of simulated air combat. All subjects progressed through all three eras during the course of the experiment.

The first era was World War II. Subjects flew either an American P-51 Mustang or a German FW-190. This era was used to train the subjects and to provide an assessment of their initial level of skill. The amount of time each subject spent in this era varied depending upon their previous pilot or simulation experience. Subjects that were completely inexperienced performed extra training sessions to learn aircraft control before beginning the air combat trials. This era typically required between 7 and 15 sessions to complete. Each session required approximately one hour to complete.

The second era was the Korean War. Subjects flew either an American F-86 Sabre or a Chinese Mig-15. This era was used to further develop the subject's skill. Sessions of the Korean War simulations were alternated with sessions of performing the subsidiary training task (i.e., either *Space Fortress* or *IQ Builder* depending on the subject's group). There were six Korean War simulation sessions for each subject. There were also 10 one-hour sessions devoted to the subsidiary task training (either *Space Fortress* or *IQ Builder*).

The third era was the Vietnam War. Subjects flew either an American F-4 Phantom II or a North Vietnamese Mig-21. This era was the final test of the subjects' abilities. There were seven sessions during this era. This era was also somewhat more complex than the previous eras due to the addition of missiles, an aircraft radar display, an aircraft radar warning display, and missile counter-measures (chaff and flares).

Experimental Task Manipulations

Regardless of era, there were two manipulations of the simulation task trials. First, the trials were classified as either easy or difficult. Easy trials usually had simpler mission goals, fewer opponents, and/or opponents with less capable aircraft or less skill. This was intended to be a straightforward manipulation of workload.

The second manipulation was the SA Tools manipulation. When SA Tools were on the subjects saw a simple Tactical Situation Display (TSD) presented in the upper left-hand corner of the monitor, and information about any selected aircraft presented in the upper right-hand corner. As the name implies, the SA Tool manipulation was intended to be a direct manipulation of the SA information provided to the pilot by the displays. The availability of such tactical information is among the most commonly suggested methods of improving SA in operational settings (Ref 7).

Dependent Measures

On each simulation trial, the following performance measures were recorded: number of bullets fired, bullet hit ratio (number of bullets that hit the opponent/number of bullets fired), number of opponents shot down (a.k.a. number of kills), and mission accomplishment (accomplished or failed).

On selected trials the three-dimensional version of the Situation Awareness Rating Technique (SART) was used (Ref 8). The subject marked his/her responses on a rating form held on a clipboard. Some trials were also used to collect a memory probe metric of situational awareness. This procedure was designed to follow the Situation Awareness Global Assessment Technique (SAGAT) procedure as closely as possible (Ref 9, Ref 10). The memory probe trial would start like any other trial, but would be stopped unexpectedly after a preselected random time interval had passed. The memory probe questions were on sheets of paper held on a clipboard.

RESULTS

All of the data were analyzed with Group (SF vs. IQ) x Era (WWII vs. VN) x Difficulty (Easy vs. Hard) x SA Tools (On vs. Off) analyses of variance (ANOVAs). In evaluating the effectiveness of the attentional control training, any interactions involving both Group and Era were considered especially important.

Performance Data

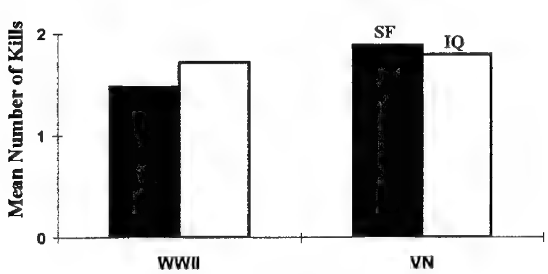
Era of combat had a significant effect on the number of bullets fired ($F(1, 30) = 283.536, p < 0.001$) and the number of kills ($F(1, 30) = 15.596, p < 0.001$). More bullets were fired in the World War II era (Mean = 606) than in the Vietnam era (84). However, there were more kills in the Vietnam era (Mean = 1.84) than in the World War II era (1.60). The difference in number of bullets reflects the different weapons technology simulated for the two era. The World War II aircraft were more dependent upon machine guns and cannons of lesser lethality. The increased skill of the subjects, as a result of more practice, may be responsible for the higher kill rate in the Vietnam era.

Mission difficulty had a pervasive effect on all performance measures. An increase in difficulty tended to: (1) increase the numbers of bullets fired, $F(1, 30) = 10.369, p = 0.003$; (2) increase the bullet hit ratio, $F(1, 30) = 53.018, p < 0.001$; (3) increase the number of kills, $F(1, 30) = 68.084, p < 0.001$; and decrease the probability of mission accomplishment, $F(1,30) = 13.582, p = 0.001$. Most of these effects may be related to the fact that the more difficult mission typically involved dealing with more opponents than the easy missions.

The availability of SA Tools significantly increased the probability of mission accomplishment ($F(1, 30) = 4.652, p = 0.039$). Without the SA Tools, the probability of accomplishment was 0.55. With the SA Tools, the probability of accomplishment increased to 0.61.

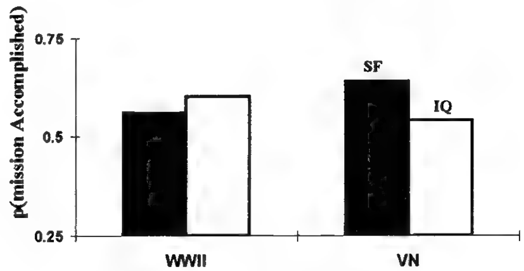
There were two Group x Era Interactions detected in the performance data. As shown in Figure 1, the SF group's mean number of kills increased more from World War II to Vietnam than did the IQ group, $F(1, 30) = 6.820, p = 0.014$.

Figure 1 - Group x Era Interaction in Number of Kills



As shown in Figure 2, the SF group's probability of mission accomplishment increased in going from World War II to Vietnam, but the IQ group's probability of mission accomplishment decreased, $F(1, 30) = 5.837, p = 0.022$.

Figure 2 - Group x Era Interaction in p(Mission Accomplished)



Both Group x Era interactions are consistent with the hypothesis that exposure to the attentional control training aided the SF group.

SART Ratings

The three dimensions of the SART (i.e., Demand, Supply, and Understanding) were analyzed separately. Mission Difficulty increased the average Demand rating from 3.9 in the easy missions to 4.4 in the hard missions, $F(1, 30) = 29.097, p < 0.001$. The availability of the SA Tools increased the mean rating of Understanding (Mean = 5.1) compared to the conditions without the SA Tools (4.9), $F(1, 30) = 4.205, p = 0.049$.

Memory Probe SA Measure

The subjects' accuracy in answering the memory probe questions was higher in the Vietnam era than in the World War II era ($F(1, 30) = 12.834, p = 0.001$). This may be a result of the increased experience subjects had by the time they performed the Vietnam missions. Memory probe accuracy was also greater if the SA Tools were on rather than off ($F(1, 30) = 60.856, p < 0.001$).

DISCUSSION

There are three issues that the current results bear upon: (1) Was the performance benefit of the attentional control replicated? (2) Were the SA metrics effective in detecting changes in SA? (3) Given that the attentional control effect was replicated and the SA metrics were effective, was the attentional control effect associated with an increase in SA?

Attentional Control Effect

The performance effects observed in this experiment were consistent with the previous research (Ref 4, Ref 5). The group that received the attentional control training (i.e., Group SF) did improve significantly more in the air combat task than did the placebo control group (Group

IQ). Interestingly, this effect was not detected in the more microscopic performance measures (such as number of bullets fired or bullet hit ratio), but was detected in the more global outcome measures (i.e., number of kills and probability of mission accomplishment). The effect was especially potent in the probability of mission accomplishment data, in which the SF group overcame a slight deficit relative to the IQ group in the World War II era missions to create a substantial lead in the Vietnam era missions (See Figure 2).

In short, the performance benefit of the attentional control training was replicated.

SA Metrics

The Difficulty and SA Tools manipulations provided an opportunity to test the sensitivity of the SA metrics used in this experiment. Both metrics responded appropriately to these manipulations. The SART subscales responded independently to the two manipulations. The Demand subscale reacted to the Difficulty manipulation, but not to the SA Tools manipulation. In contrast, the Understanding scale reacted to the SA Tools manipulation, but not to the Difficulty manipulation.

The accuracy of the memory probe responses was much better in the missions using SA Tools than in the missions without the SA Tools. In fact, inspection of the magnitudes of the F values associated with the three metrics that responded to the SA Tools manipulation (i.e., probability of mission accomplishment, SART Understanding scale, and memory probe accuracy) suggests that the memory probe metric was the most sensitive to the SA Tools manipulation.

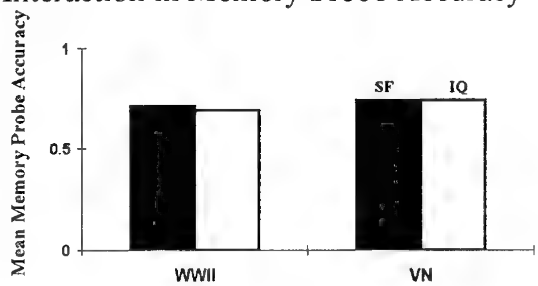
So, within the context of the experimental trial manipulations the SA metrics employed in this experiment appear to have been responsive to SA manipulations.

Attentional Control and SA

The results of this experiment provide no support for the hypothesis that the beneficial effect of the attentional control training is mediated through SA improvement. If this hypothesis were correct, then a Group x Era interaction of the same form as observed in the number of kills (Figure 1) and probability of mission accomplishment (Figure 2) would be expected. Subjective rating metrics, such as the SART, might have difficulty in detecting this between-subject effect because people may tend to rate the different conditions within the overall context of their personal experience. However, the memory probe metric would be expected to be sensitive to any beneficial effect of the attentional control training on the quality of the SA achieved by the two groups.

The memory probe results are clearly inconsistent with the hypothesis. Not only did the Group x Era interaction fail to reach statistical significance in the memory probe accuracy data ($F(1, 30) = 0.481, p = 0.493$), but the mean memory probe accuracy for the two groups is nearly identical across the two eras (See Figure 3).

Figure 3 - Lack of Group x Era Interaction in Memory Probe Accuracy



In short, the results of the present experiment support the effectiveness of attentional control training on performance enhancement, support the utility of SART and memory probes as SA metrics, and fails to support an association of the attentional control training with improved SA. It remains to be determined whether the hypothesized relationship between attentional control and SA is incorrect or whether there are aspects of SA that the metrics used in this study do not capture.

REFERENCES

1. Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting, 32,* 97-101.

2. Carroll, LA (1992, March). Desperately seeking SA. *TAC Attack (TAC SP 127-1), 32*(3), 5-6.

3. Dominguez, C. (1994, June). Can SA be Defined? In M. Vidulich, C. Dominguez, E. Vogel, & G. McMillan (Eds.), *Situation Awareness: Papers and Annotated Bibliography* (Tech. Rep. No. AL/CF-TR-1994-0085, pp. 5-15). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

4. Gopher, D., Weil, M., & Bareket, T. (1992). The transfer of skill from a computer game trainer to actual flight. *Proceedings of the Human Factors Society 36th Annual Meeting, 36,* 1285-1290.

5. Hart, S., & Battiste, V. (1992). Field test of video game transfer. *Proceedings of the Human Factors Society 36th Annual Meeting, 36,* 1291-1295.

6. Gopher, D. (1993). The skill of attentional control: Acquisition and execution of attention strategies. In D.E. Meyer & S. Kornblum (Eds.) *Attention and Performance XIV* (pp. 299-322). Cambridge, MA: MIT Press.

7. Vidulich, M.A., Dominguez, C., Vogel, E., & McMillan, G. (1994, June) (Eds.), *Situation Awareness: Papers and Annotated Bibliography* (Tech. Rep. No. AL/CF-TR-1994-0085). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

8. Taylor, R.M. (1990, April). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *AGARD-CP-478, Situational Awareness in Aerospace Operations*, 3-1 - 3-17.

9. Endsley, M.R. (1990, April). A methodology for the objective measurement of pilot situational awareness. *AGARD-CP-478, Situational Awareness in Aerospace Operations*, 1-1 - 1-9.

10. Endsley, M.R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology, 3,* 157-168.

Agent Based Multimedia Dialogues for Reduced Workload and Increased Situational Awareness

Malcolm James Cook and Leona Elder,
University of Abertay,
Bell Street,
Dundee, DD1 1HG
United Kingdom
contact: m.cook@abertay-dundee.ac.uk
Tel: (UK) 1382 308464
Fx: (UK) 1382 308877

Abstract

Situational awareness is the term used to describe the crew or pilot's awareness of operational conditions and contingencies and it has been implicated as a contributory factor in incidents and accidents in the air (Helmreich and Foushee, 1993). Analysis of incidents reveals that pilots often fail to perceive a problem existed or that the significance of the cues with respect to the safety of the flight was overlooked (Orasanu, 1993). Presenting information in a suitable way can increase the probability that a specific event is noticed by the pilot. The method of presentation, however, can not ensure that the knowledge of the event is maintained or that its significance in situational terms is recognised. This paper examines evidence from the literature on vigilance tasks (sustained attention tasks) and dual task experiments which indicates that limits of short term memory, switching attention or time taken to access long-term memory can exacerbate the problem of maintaining situational awareness. Evidence from two experiments are discussed which indicate that the bottleneck for processing information in complex tasks may be related to attention switching and limits on short-term (working memory).

Introduction

It seems obvious to suggest that the crew or pilot's effectiveness depends on situational awareness. It is equally obvious that factors which interfere with maintaining good situational awareness can result in impaired decision-making and an increased number of accidents. Workload is a factor which can promote and degrade situational awareness. Workload is the term used to describe the processing capacity devoted to a task or tasks by an operator (Damos, 1991). If workload is too low the pilot may be required to maintain attention and scan instruments for current information. Low workload may result in under-arousal and this in turn may decrease the quality of information processing as the pilot's state falls to the under aroused portion of the Yerkes-Dodson curve. High workload may result in over-arousal and this may be combined with an information rate from multiple sources that exceeds the pilot's capacity.

Workload is both related to situational awareness and is influenced by it (Adams, Tenney and Pew, 1991). A central factor influencing workload management is the ability to schedule competing tasks for processing, so that the most task critical at that time is accomplished first. The effective instantaneous prioritization of tasks and anticipatory behaviours of the pilot depend on good situational awareness. Seeking to change situational awareness in the absence of reductions in workload may be ineffective because the scheduling of a complex and demanding set of tasks may actively interfere with situational awareness. Scheduling is a task in itself requiring processing resources which may be required for adequate processing of core tasks.

The non-linear relationship between workload and the quality of processing of information is very important. Whether one accepts the levels of processing argument for the retention of material in memory (Craik and Lockhart, 1972), or not, the quality of processing which information receives clearly has an effect on recall and recognition performance. The probability of recalling appropriate information, in turn, relates to the quality of situational awareness because better recall means a clearer picture of the true facts. Situational awareness in turn can affect the quality of decision making.

The problem with this line of argument is the assumption that the pilot is a passive receiver of information. Evidence from pilots and from other experts carrying out decision making tasks suggests that perception, encoding and retention of information are active processes. Thus, expert pilots often disregard information and in the process actively degrade their own decision making. The rejection of or failure to use all the information available has been found to be particularly damaging when situations experienced or problems encountered are non-routine (see Chapter 8, Huey and Wickens (1993)). One of the factors which clearly contributes to decision-making is situational awareness and that is why it is such an important area of study. The availability of information, as a consequence of situational awareness, does not in itself guarantee that a

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

logical conclusion will be generated. Poor situational awareness can co-occur with too much, or too little, information and limited time to thoroughly process the information (Morgan, Herschler, Weiner and Salas, 1993).

Information Salience and Situational Awareness

If one accepts the active nature of the perceptual, encoding and retention processes without reservation one might erroneously conclude that manipulating the way in which information is presented will have little effect on the quality of decision making. It can, however, be argued that presentation can influence the pilot's decision making. It has been argued that computer-assisted detection systems might be constructed in a way that increases the instantaneous probability and time-integrated probability of detecting target events (Wickens and Huey, 1993). For instance a limited amount of experimental evidence supports the efficacy of enhanced audio-visual displays but care must be taken in considering the theoretical and experimental evidence as enhanced information delivery can reduce or increase performance (Morgan, Herschler, Weiner and Salas, 1993). The key factor that seems to play a part in the disruptive effect of enhanced information delivery is the attention capturing properties of the enhanced auditory or visual signal and the temporal location of the signal in relation to the pilot's processing of other events. Indeed, the paucity of evidence demands more experimental examination of this area before enhanced displays and information delivery are used to facilitate information transfer during high workload flight segments where misdirected attention may accrue high costs.

In considering cognitive modelling of fighter aircraft process control Amalberti and Deblon (1992) suggested that the pilot would only benefit from an on-board intelligent assistant if the behaviour of the assistant was well understood. They reasoned that part of this understanding between the pilot and an intelligent assistant rested on adequate communication. By seeking to influence the registration of and maintenance of information through multimedia displays the pilot's understanding of the assistant and their situational awareness can both be improved giving better outcomes in terms of action. The provision of the same information in different modes of communication means that the pilot's load can be balanced between tasks. For example, during tasks that present information visually the pilot can accrue an awareness of other significant events in the auditory mode of presentation. During high demand the required tasks can be re-designed to provide signals across modalities to maximise the bandwidth of information or to redundantly encode important signals.

Judicious use of enhanced information delivery does seem to offer improvements in decision making and that may be attributable to the salience of the information delivered. Wickens and Flach (1988) have suggested a model of decision making and they have listed a number of heuristics and biases in the processes involved. One of the biases is known as the *Salience Bias*. Salience bias describes the decision maker's tendency to focus on the most salient cues and use that information in decision making. If intelligent assistant systems can identify critical information related to the current flight status and the information can be presented in a salient form, it should influence the pilot's actions. In effect one would modify the pilot's situational awareness.

Such considerations relate directly to accidents such as that in which an Air Florida 737 crashed in bad weather shortly after takeoff. The pilot flying the Air Florida Boeing 737 was unaware of a significant fault in the system measuring engine pressure ratio (EPR). The EPR was actually lower than indicated on some of the instruments and the pilot failed to notice the position of the throttle. The plane stalled and crashed. Delivery of a salient cue at an appropriate time might help prevent accidents such as this by challenging the pilot's model of the world. Boy (1991) has suggested that the pilot's behaviour illustrates two types of error. First, the pilot fixated on the information related to the icing of the plane because of delays to take-off and the weather. Second, the pilot improvised and employed non-standard actions to de-ice the plane prior to take off. The first type of error could arguably be challenged by appropriate cueing. The First Officer did on a number of occasions, indicate a possible problem to the Captain of the plane. The Captain clearly ignored the First Officer but the response to an intelligent assistant urging action might have been different. Analysis of accidents reveals that failure to perceive a problem or a failure to recognise the significance of cues is a major factor in the generation of aerospace accidents and incidents (Freeman and Simmon, 1991).

The argument put forward here is an extension of that put forward by Adams, Tenney and Pew (1991). Adams et al. (1991) suggest that supervisor-operators' performance can be improved by improving the available information and increasing the likelihood that the operator has the necessary mental models in memory. Improving the salience of information and structuring the delivery of information in time will ensure that this is more likely to be the case. An area that needs examination is the flexibility of user models in low and high workload situations or under low, moderate and high states of arousal. It may be that earlier and more salient warnings under relatively low workload may be more successful in promoting good situational awareness. Or, it may be equally difficult to change the pilot's model of the world in high and low workload situations.

Confirmation Bias and Situational Awareness

Wickens and Flach (1988) suggest that the other heuristics and biases that operate within the decision making process may actively resist a change to the pilots model of the world. For example, the *confirmation bias* describes the tendency for pilots to look for information that confirms or supports the current hypothesis and the tendency to ignore or reject information that runs counter to the current hypothesis. While these tendencies to support biased judgement and reject information are not restricted to real-time response high workload situations (Camerer and Johnson, 1991; Evans, 1992) there is little evidence of the effects of workload on the strength of the bias even though there are intuitive expectations of the outcome.

Accepting that one can influence behaviour by keeping information salient one might consider the factors that control the successful transfer of information from the sensory register to working memory and the subsequent processing that takes place. The difficulty with this analysis is the role of long term memory and the information it affords for the interpretation of events. Clearly people use information from long-term memory to assist in the processing of current information. The extent to which tasks are completed successfully may to some degree depend on information in long term memory.

Vigilance, Dual-Task Performance and Situation Awareness in the Cockpit

The experiments examined in this paper use vigilance tasks and dual-tasks to explore the processing of information. The aim was to identify the ways in which information presentation might be enhanced to improve the salience of the cue by promoting more effective processing. It has been suggested that it is hard to think of ways to improve an individual's performance on a dichotic listening task, or reasons, other than purely theoretical, for wanting to try (Adams, Tenney and Pew, 1991). It is argued that more effective processing of information may be the key to improving situational awareness because it results in better recall from memory.

The justification for using evidence from adapted vigilance tasks is the sustained nature of the information processing tasks in the cockpit. Authors like Adams, Tenney and Pew (1991) have accepted that there is a relationship between workload and situational awareness. Textbooks, reviewing Dual-Task Performance like that edited by Diane Damos (1991), rarely or never make reference to the effects of workload on situational awareness and the interaction between the two measures of performance.

Adams et al. (1991) have suggested that interruptions are a particular problem in multi-task situations and that the reception of the arriving information introduces an additional disruptive element of workload. Further they suggest that with proper timing, the disruptive effects of these interruptions can be minimised. Clearly this management of incoming information must be a compromise that does not prevent the pilot from requesting further information because this may introduce a systematic bias in the information received. The benefits of temporal management of signals by intelligent agents fielding information in specific domains, is the reduction of workload at critical points in flight and the improved situational awareness within the relevant domains. Clearly this type of system must be interactive if it is to behave intelligently because it must have knowledge of the pilots' goals, actions, mental models, and attentional resources as well as the temporal requirements, deadlines, priorities and interdependencies among tasks.

First Experimental Series

In simple terms, the key factors determining situational awareness, other than prior experience, are the availability of attentional resources and free capacity in working memory. In the first series of experiments, extended watch visual detection tasks were varied in terms of the demands the tasks put on memory. Three conditions were examined in which subjects were asked to detect one of three types of event and the recurrent memory load was increased across the tasks.

The basic visual stimulus shown on a VDU to subjects was a five by five grid in which sixteen positions were filled with coloured squares. Twelve of the squares were blue and four were red. The display of squares flashed on for 400ms and the inter-stimulus interval was 600ms. A thousand trials were displayed and forty five of these contained a significant event which the subject had been asked to detect. The first condition involved a set pattern repeated across the non-event trials but all the squares changed to red on trials that were to be treated as events. The second condition involved a repeated pattern on the non-event trials and in the event trials one of the randomly positioned red squares changed position. The third condition involved a random pattern of square distribution on non-event trials and the event the subjects detected was a complete repeat of a pattern from one presentation to another.

It was anticipated that in the first condition the memory load would be very light and subjects would make an involuntary shift in attention to appearance of the colour change of the squares during a event. The second condition

involved a moderate memory load and pilot work indicated that the change in square position would require some degree of sustained attention to ensure detection. The third condition was clearly very demanding and without chunking (Miller, 1956) of information in memory the pattern would exceed the capacity of working memory. The constant changes to the pattern would limit the involvement of long-term memory. It was expected that performance would deteriorate to the greatest extent in the third condition. The third and second conditions were repeated in a smaller number of subjects with auditory cues to indicate the nature of the events presented i.e. to redundantly code target events to examine the effects of redundant multimedia cueing.

Results

Mean event detection for the task in which all squares became red was 44.6 (N=14, SD=0.50) and the mean number of false alarms was 0.93. The mean event detection for the second task, in which one red square moved position, was 40.4 (N=14, SD=3.5) and the number of false alarms was 14.1 (N=14, SD=19.53). The mean event detection was lowest at 34.0 (N=19, SD=7.6) in the third condition where a repeat of a previous pattern was the event the subject was asked to detect. The false alarm rate was also very high in the third condition (18.6, N=19, SD=10.54). Where auditory cues were used to indicate the significance of events subjects' performance improved with the number of false alarms falling and the number of events detected rising as compared to the equivalent dual task conditions. The number of events detected for the most difficult square displays was 42.8 (N=8, SD=3.4) and the number of false alarms 3.3 (N=8, SD=3.6).

Second Experimental Series

In the second set of experiments eight conditions were tested in a 2 by 4 between subjects design. The first factor was the type of display shown to subjects and the second factor was the type of task(s) subjects were asked to carry out. The type of display shown to subjects was basically similar to that in the first and second conditions of the first experimental series. The first condition was treated as a low demand task because the red squares would elicit an involuntary shift in attention. The second condition in which a single square moved position to indicate an event was treated as a high demand task. The only additional feature in both the low and high demand displays was an arrow that changed orientation between each display. Three of the four orientations the arrow assumed (vertical, left and right facing) were frequent and the fourth orientation (facing down) occurred on only forty five of the thousand events in the trial. Subjects in the low and high demand conditions were given one of four instructions. The subjects could be asked to do a single task or a dual task. If they were asked to carry out a single task, they were asked to carry out the vigilance task involving either squares or the detection of the downward facing arrow. If they were asked to attempt both tasks they were asked to ensure that one task of the two was accomplished to the best of their ability and any spare capacity given to the secondary task. It was stressed that high scores on both were preferable but they should maximise their effort on the task designated as the primary task.

Results

The results from the single task conditions in both low and high demand variants were almost identical to those from the previous set of experiments (see Tables 1 and 2).

The performance in the dual task conditions exhibited a reduced event detection performance and in most cases a substantial increase in false alarms.

With auditory cueing the event detection performance in the dual task condition, where the subjects detected a moving square, approached saturation (43.0, N=8, SD=3.1) and performance for the orientation task was also close to saturation (43.3, N=8, SD=3.6). The false alarm rate with auditory cueing of the orientation task was low (3.25,N=8, SD=2.9). The false alarm rate for the detection of the moving square in the dual task condition with auditory cueing was equally low (2.0, N=8, SD=0.82).

Discussion

There are two significant effects of redundantly coded information in this paper. First, the increase in event detection performance when redundantly coded target events are presented as compared to the performance where events are cued in a single modality. Secondly, the decrease in false alarm rates when redundantly coded information is presented. This reduction in false alarm rate is marked when the memory load, of the single modality variant, is very high. The increase in event detection performance and decrease in false alarm rate in both tasks during a dual task situation suggests that the effective memory and attentional load is reduced substantially.

Reducing the number of false alarms in response signals in the environment and increasing the probability of detecting signals seems like a reasonable aim. The underlying goal is to increase situational awareness by artificially promoting the salience of critical information and to decrease workload in the process allowing more effective processing of all information. Situational awareness and workload are closely tied together particular when signal detection performance is poor. As the number of false alarms a pilot responds to increases the quality time available for

processing other information decreases. It seems that at high memory loads operators are particular prone to false alarms. This seems to depend on an awareness of increasing failure to detect signals and it could be a behavioural response to that decrease in performance. Whatever the proximate cause of the false alarms overall performance in simultaneous tasks could be improved by eliminating unnecessary responses to non-signals. Clear and simple signalling of appropriate events at appropriate times in flight would reduce the processing of incoming information. The simplicity and clarity of the information presented might be further increased by scheduling events after pre-processing and under the control of autonomous agents. Transferring the interpretative burden an intelligent assistant would in turn reduce the memory load and in turn decrease the possibility of costly false alarms.

Further data collection for the findings reported is underway and the pattern of dual-task interactions will be studied in detail. A temporal analysis of performance is clearly required to determine what factors, if any, are associated with misses and false alarms. It may be that dual-modality tasks can in some circumstances generate more misses or more false alarms when transitory increases in signal rate occur. Subjects did report that transitory increases in event rate were subjectively demanding and that this may be worse with dual-modality delivery. It may be the case that divided attention across modalities improves performance as one would expect from reviews of the literature and models derived from that literature (cf. Stokes, Wickens and Kite, 1990). However, during a rapid succession of events there may be inertia developing in or inherent in the system for switching attention between modalities. There are certainly reports of individual differences in the ability to switch attention (Gopher and Kahneman, 1971; Kahneman D., Ben-Ishai R. and Lotan M. , 1973). Clearly there may be direct interactions with the individual differences and signal event rate across modalities. This suggests that temporal factors in information delivery need further investigation to establish the nature of these putative interactions.

Subjects reported that they seemed to more aware of local event probabilities during auditory cueing and this may suggest that auditory delivery can improve situational awareness. The salience of auditory cues in dual tasks does not conform with the expectations derived from the phenomenon of visual dominance identified by Wickens (1993) but this finding clearly needs further systematic investigation. It is possible that auditory encoding of processed events degrades more slowly and it may be this, and not attentional factors, that contributes to the salience of auditory events a point noted by Wickens (1993).

Work is currently in progress to establish if the workload produced in the different conditions affects a planning or scheduling task which is related or unrelated to the events detected. Establishing that events are detected is clearly distinct from the use of the event related information in decision making. The results indicate that mixed modality presentation of information could cause problems at high workload and further work is required to establish if interactions exist with the probability of an incorrect signal. If behaviour is adversely affected by information presentation in multi-modal signals and the performance decrement is increased futher by small percentages of incorrect signals then the use of direct voice input and output systems should be restricted at times of high workload.

Situational awareness can not be explained simply in terms of quality of information processing of incoming stimuli but many recognise that failure to acquire important information is a major contributor to accidents and incidents that associated with poor situational awareness (Woods, Johannesen, Cook, and Sarter, 1994).

References

Adams M.J., Tenney Y.J. and Pew R.W. (1991) Strategic Workload and the Cognitive Management of Advanced Multi-Task Systems, Crew System Ergonomics Information Analysis Center, Wright-Patterson Air Force Base, Ohio, USA.

Amalberti R. and Deblon F. (1992) Cognitive modelling of fighter aircraft process control : A step towards an intelligent on-board assistant system, International Journal of Man-Machine Studies, 36, 639-671.

Boy G. (1991) Intelligent Assistant Systems, Academic Press, London.

Camerer C.F. and Johnson E.J. (1991) The process-performance paradox in expert judgement : How experts know so much and predict so badly. in Towards a General Theory of Experise: Prospects and Limits, K. Anders Ericsson and J. Smith Eds., Cambridge University Press, Cambridge, England.

Craik F.I.M. and Lockhart R.S. (1972) Levels of Processing : A Framework for Memory Research, Journal of Verbal Learning and Verbal Behaviour, 11, 671-684.

Damos D. (1991) Multiple-Task Performance, Taylor-Francis, Bristol, England.

Evans J. St. B.T. (1992) Bias in thinking and judgement. in Advances in the Psychology of Thinking Volume One, Harvester Wheatsheaf, Hemel Hempstead, Herts., England.

Freeman C. and Simmon D.A. (1991) Taxonomy of crew resource management : Information processing domain, Proceedings of the Sixth International Symposium on Aviation Psychology, 391-397, Columbus: The Ohio State University, USA.

Gopher D. and Kahneman D. (1971) Individual differences in attentionand the prediction of flight criteria, Perceptual and Motor Skills, 33, 1335-1342.

Kahneman D., Ben-Ishai R. and Lotan M. (1973) Relation of a test of attention to road accidents, Journal of Applied Psychology, 58, 113-115.

Helmreich R.L. and Foushee H.C. (1993) Why Crew Resource Management ? Empirical and Theoretical Bases of Human Factors Training in Aviation in Cockpit Resources Management, E.L. Weiner, B.G. Kanki and R.L. Helmreich Eds., Academic Press, London.

Huey B.M. and Wickens C.D. (1993) Workload Transition : Implications for Individual and Team Performance, National Academy Press, Washington DC, USA.

Miller G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing infomration, Psychological Review, 63, 81-97.

Morgan Jr. B.B., Herschler D.A., Weiner E.L. and Salas E. (1993) Implications of Automation Technology for Aircrew Coordination and Performance in Human/Technology Interaction in Complex Systems, Vol. 6, pp 105-136, W.B. Rouse Ed., JAI Press Inc., Middlesex England.

Orasnu J.M. (1993) Decision-making in the Cockpit in Cockpit Resources Management, E.L. Weiner, B.G. Kanki and R.L. Helmreich Eds., Academic Press, London.

Stokes A., Wickens C. and Kite K. (1990) Display Technology : Human Factors Concepts, Society of Automotive Engineers, USA.

Wickens C.D. and Flach J. (1988) Human Information Processing. in Human Factors in Aviation, New York, Academic Press, London.

Woods D.D., Johannesen L.J., Cook R.I. and Sarter N.B. (1994) Behind Human Error : Cognitive Systems, Computers, and Hindsight, CSERIAC Report 94-01, CSREIAC, Wright-Patterson Airforce Base, Ohio, USA.

Table 1 : Results For Single Task Conditions Used As Controls For Dual Tasks.

Squares Task (Low Demand)		Orientation Task	
Event Detections	False Alarms	Event Detections	False Alarms
44.3 (N=9, SD=1.1)	0.2 (N=9, SD=0.67)	44.2 (N=9, SD=0.97)	0.56 (N=9, SD=0.88)
Squares Task (High Demand)		Orientation Task	
Event Detections	False Alarms	Event Detections	False Alarms
42.6 (N=9, SD=1.8)	2.4 (N=9, SD=2.1)	43.8 (N=10, SD=1.5)	1.0 (N=10, SD=1.1)

Table 2 : Results For Dual Task Conditions

PRIMARY TASK		SECONDARY TASK	
Squares Task (Low Demand)		Squares Task (Low Demand)	
Event Detections	False Alarms	Event Detections	False Alarms
43.7 (N=10, SD=1.2)	1.0 (N=10, SD=0.88)	43.2 (N=11, SD=2.5)	4.18 (N=11, SD=7.7)
Squares Task (High Demand)		Squares Task (High Demand)	
Event Detections	False Alarms	Event Detections	False Alarms
37.6 (N=11, SD=5.3)	6.6 (N=11, SD=5.26)	39.5 (N=10, SD=4.6)	6.8 (N=10, SD=6.9)
Orientation Task (with low demand squares task)		Orientation Task (with low demand squares task)	
Event Detections	False Alarms	Event Detections	False Alarms
42.8 (N=10, SD=2.6)	2.2 (N=10, SD=2.4)	43.1 (N=11, SD=1.7)	1.5 (N=11, SD=1.57)
Orientation Task (with high demand squares task)		Orientation Task (with high demand squares task)	
Event Detections	False Alarms	Event Detections	False Alarms
33.8 (N=11, SD=9.2)	4.6 (N=11, SD=2.8)	38.4 (N=10, SD=4.0)	5.1 (N=10, SD=5.0)

Figure 1) Typical Display from dual task condition

Lightsquares were red and dark squares were blue in actual display shown. The cross was black and the background was a light grey.

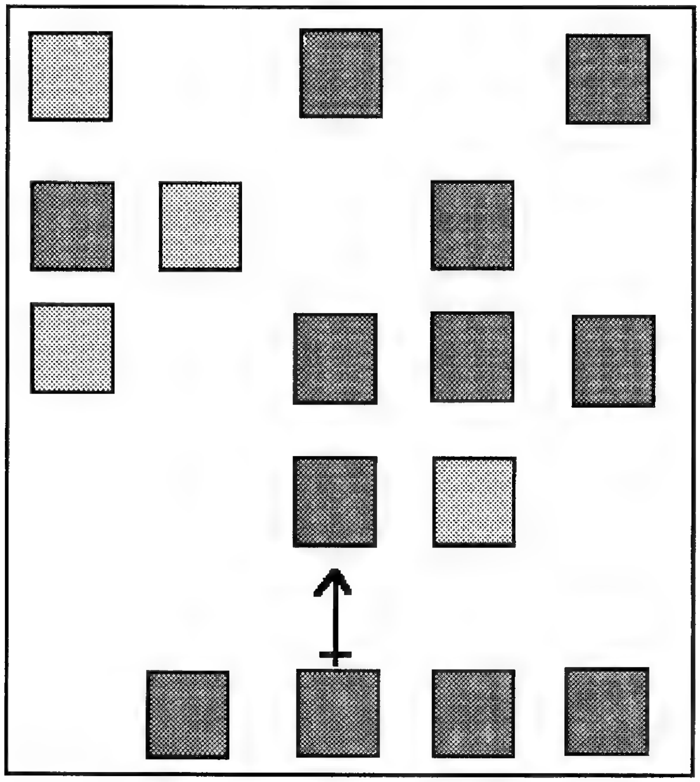
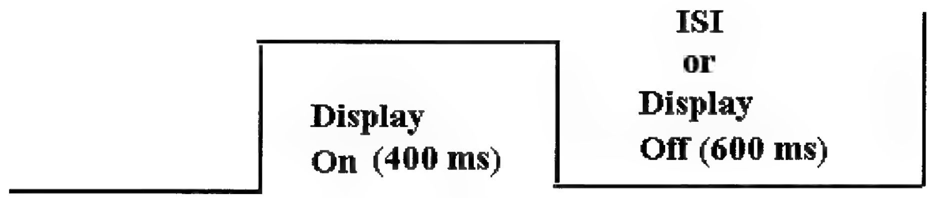


Figure 2)Temporal display details



The subject would see a display with an event and they would respond within a second after display onset.

The display details applied to all experiments in both the first and second experimental series.

Use of Multiship Simulation as a Tool for Measuring and Training Situation Awareness

Wayne L. Waag
Aircrew Training Research Division
Armstrong Laboratory
6001 S. Power Rd.
Mesa, AZ 85206-0904
USA

Michael R. Houck, David A. Greschke, and William B. Raspotnik
University of Dayton Research Institute
PO Box 2020
Higley, AZ 85236
USA

SUMMARY

This paper presents the findings of a research investigation that explored the use of networked multiship simulation as a tool for measuring and training situation awareness (SA). The Division's research simulation facility was used which permitted two F-15s to fly against a suite of manned and unmanned adversaries in a realistic combat environment. Controller support was provided using a long-haul network linked to an AWACS simulation located at Brooks AFB, TX. A week-long evaluation syllabus was designed consisting of 9 sorties with 4 engagements per sortie. A building block approach was taken so that scenarios increased in difficulty over the week. Sixty-three mission ready F-15 pilots participated in the study. Performance ratings of SA were gathered using two trained observers. Additionally, mission outcome, network communications, video recordings, and eye movement data were gathered. As expected, SA, as measured in the simulation environment, was found to be positively correlated with ratings of SA previously obtained from the pilot's home squadron. Performance in the simulation was found to improve for identical engagements flown early and late in the syllabus. Positive opinions were expressed by study participants regarding the potential value of multiship simulation for training SA skills. Areas of greatest payoff appear to be the training of flight resource management and decision-making skills. It was concluded that multiship simulation can be an effective tool for both measuring and training SA.

1 INTRODUCTION

Study Background. In 1991, the US Air Force Chief of Staff posed a series of questions concerning SA that led to the present investigation. First of all, what is SA? Can it be objectively measured? Is SA learned or does it represent a basic ability or characteristic that some pilots have and others do not? From a research standpoint, these questions translate into issues of measurement, selection, and training. The Armstrong Laboratory was subsequently tasked with providing research answers to these questions. A research investigation was initiated that had three goals: first, to develop and validate tools for reliably measuring SA; second, to identify basic cognitive and psychomotor abilities that are associated with pilots judged to have good SA; and third, to determine if SA can be learned, and if so, to identify areas where cost-effective training tools might be developed and employed. An overview of the investigation can be found in this report in the paper by McMillan, Bushman, and Judge (1).

The general approach was to first develop criterion measures of SA based upon performance ratings collected within an operational flying environment. The results of this part of the study can also be found in this report in the paper by Waag and Houck (2). These measures were necessary for two reasons. First, they would serve as criterion measures against which to validate a battery of basic ability tests considered relevant to SA, thereby addressing the question of basic human abilities. The results of this part of the study can also be found in this report in the paper by Carretta and Ree (3). Second, these measures would serve as a means of selecting a sample of pilots who would participate in a simulation phase of the effort. During that phase, simulated air combat mission scenarios were developed for assessing SA and objective measures of performance gathered in an attempt to determine those characteristics that distinguish pilots with good SA. These data would be used to identify areas where training tools might be developed. This paper presents the results of only the third phase of the program, namely the use of simulation as a tool for measuring and training SA.

The Measurement of SA. In response to the question, "what is it?" a working group at the Air Staff produced the following operator's definition of SA: "a pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception (4)." While other definitions of SA within the literature focus primarily on processes underlying the assessment and resulting knowledge of the situation (5,6), our working definition also included forecasting, decision making, and task execution. From an operational Air Force perspective, SA is more than simply knowledge and understanding of the environment.

The development and validation of measurement tools is described in detail in the paper by Waag and Houck (2). To briefly summarize, it was first necessary to identify and describe critical behavioral indicators of the fighter pilot's ability to maintain good SA and successfully complete his mission. Previously, Houck, Whittaker, and Kendall (7) conducted a cognitive task analysis of a typical F-15 air combat mission. The resulting analysis identified the significant types of decisions required of the flight members, the information required for making these decisions, and the

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

observable activities the flight members performed to acquire this information. The results were further analyzed by an experienced fighter pilot to identify behavioral indicators considered most essential to SA. This subject matter expert (SME) emphasized that these behavioral indicators must be observable in the context of day-to-day squadron training activities and subject to evaluation by fighter pilots both in terms of their own performance and that of others. As a result of this analysis, 24 behavioral indicators organized in seven categories were identified and are shown in Table 1.

TABLE 1. ITEMS AND CATEGORIES USED IN SARS

1. TACTICAL GAME PLAN
    - Developing plan
    - Executing plan
    - Adjusting plan on-the-fly
2. SYSTEM OPERATION
    - Radar
    - Tactical electronic warfare system
    - Overall weapons system proficiency
3. COMMUNICATION
    - Quality (brevity, accuracy, timeliness)
    - Ability to effectively use information
4. INFORMATION INTERPRETATION
    - Interpreting vertical situation display
    - Interpreting threat warning system
    - Ability to use controller information
    - Integrating overall information
    - Radar sorting
    - Analyzing engagement geometry
    - Threat prioritization
5. TACTICAL EMPLOYMENT-BVR
    - Targeting decisions
    - Fire-point selection
6. TACTICAL EMPLOYMENT-VISUAL
    - Maintain track of bogeys/friendlies
    - Threat evaluation
    - Weapons employment
7. TACTICAL EMPLOYMENT-GENERAL
    - Assessing offensiveness/defensiveness
    - Lookout
    - Defensive reaction
    - Mutual support

Based principally upon these behavioral indicators, a number of SA Rating Scales (SARS) were developed to measure SA in operational units. They were administered to 239 mission-ready F-15 pilots from 11 operational squadrons. From the SARS, a composite measure of SA was derived and found to be highly related to previous flight experience and current flight qualification (2). These measures were used for two purposes. First, they served as a criterion measure against which to validate a battery of basic ability tests considered relevant to SA, thereby addressing the question of basic human abilities (the second goal of the study). The Situation Awareness Assessment Battery (SAAB), consisting of 24 computer-based tests of basic cognitive and psychomotor abilities (3), was also administered to the same sample of pilots at their home units.

Second, these measures served as a means of selecting a sample of pilots who participated in a simulation phase of the effort, in which performance was observed under realistic combat conditions. During this phase, simulated air combat mission scenarios were developed for assessing SA and a variety of performance measures gathered in an attempt to determine whether SA could be measured in a simulation environment. Moreover, an attempt was made to examine the potential of this type of simulation for training critical SA skills. This paper presents some preliminary findings of the data gathered from a simulated air combat environment.

2 METHOD

Subjects. A total of 40 mission-ready (MR) F-15 pilots, who were flight lead qualified served as subjects. An additional 23 MR F-15 pilots served as wingmen throughout the data collection which began in Mar 93 and was completed in Jan 94.

Simulation System. The Armstrong Laboratory multiship simulation facility (MULTIRAD) located at Williams Air Force Base (WAFB), Arizona (now Williams Gateway Airport, Mesa, AZ) was used. The major components of the simulation system are shown in Figure 1. These components

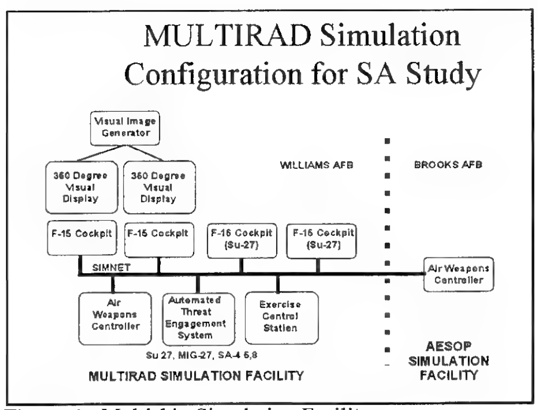


Figure 1. Multiship Simulation Facility

represent independent subsystems operating as part of a secure distributed simulation network. This local area network was connected to the air weapons controller simulator (AESOP) at Brooks Air Force Base (BAFB), TX by a dedicated T-1 telephone line. Additional details concerning the basic simulation architecture and components are available in Gehl, Rogers, Miller, and Rakolta (8) and Platt and Crane (9).

The manned flight simulators consisted of two F-15C simulators and two F-16 simulators. The F-15C simulators had high fidelity aerodynamic, engine, avionics, radio, sensor, and weapons simulations. Each F-15C simulator was equipped with an out-the-window visual display system covering approximately 360 deg horizontal by 200 deg vertical. The external visual scene was created using computer-generated imagery. The manned F-16 simulators had less fidelity and played the role of enemy aircraft in conjunction with computer-controlled adversaries. The visual and electronic signatures of these F-16 simulators were modified so that they appeared as the appropriate threat aircraft. Each F-16 simulator was equipped with a single channel of out-the-window visual imagery covering approximately 45 deg horizontal by 45 deg vertical.

A manned air weapons controller (AWC) provided the F-15C pilots with appropriate threat information and warnings. Depending upon the availability of qualified AWCs and equipment status, the AWC was either located at WAFB or BAFB. In either case, the AWC had a realistic simulation of the appropriate AWC console and communicated with the F-15C pilots by radio.

The exercise control system (ECS) consisted of a central console with the hardware and software necessary to create, start, observe, record, and stop the simulated air combat sorties. The SMEs who served as test directors and observers viewed monitors that provided a real-time view of each sortie. These monitors provided: 1) a plan view display of all the participants in each engagement along with status information; 2) the instrument panel of each F-15C cockpit which included the radar, radar warning receiver, and armament displays; and 3) the forward channel of out-the-window video for each F-15C cockpit. The plan view display, instrument panel displays, and radio communication were also recorded to video tape for mission debrief and further data analysis. In addition, the ECS included a data logger that recorded all the network communication protocols between simulators.

Ground threats, as well as additional threat and friendly aircraft, were provided by a computer- based automated threat engagement system (10). The ground threat portion of the automated threat engagement system (ATES) provided command and control functions (e.g., early warning radars and target assignment) and simulation of directed and autonomous surface-to-air missile batteries and anti-aircraft artillery with their radars. The aircraft portion of the ATES provided computer controlled air interceptors as well as formations of air-to-ground bombers. In addition, the ATES provided four computer controlled F-16s which were escorted by the manned F-15Cs during offensive counter air sorties.

Scenario Design. The primary approach taken toward the measurement of SA was through scenario manipulation and observation of subsequent performance as recommended by Tenney, Adams, Pew, Huggins, and Rogers (11). Other approaches such as the use of explicit probes (5) were considered and finally rejected due to their lack of face validity for the study participants. Since we were using mission-ready F-15 crews, it seemed essential that we provide a simulation experience as realistic as possible. A week-long SA "evaluation" exercise was constructed that consisted of 9 sorties with 4 engagements per sortie. Sorties were arranged in a building block manner. Over the week, engagements increased in complexity in terms of numbers of adversaries, enemy tactics, lethality of ground threats, AWC support, etc.

A typical engagement scenario is presented in Figure 2. This depicts a defensive counter air (DCA) mission in which the objective of the two F-15s is to defend the home airfield. In this case, the attackers consist of two bombers accompanied by two fighters. The engagement begins at 80 nautical miles (nm) separation in which the fighters are flying at 20,000 ft. and the bombers at 10,000 ft. They are laterally separated by 10 nm which makes them fairly easy to acquire on radar by the two F-15s. At 35 nm, the fighters begin a corkscrew type of maneuver in which they rapidly descend to 3500 ft. At this time, they will drop off of the F-15's radar screen. Upon completion of the maneuver, the fighters will trail the bombers as well as be at a much lower altitude. While the F-15s can easily continue tracking the bombers, it requires the crew to "predict" the actions of the fighters so that they may be quickly re-acquired on radar. At 15 nm, the bombers do a hard right turn and descend to 2500 ft. At this time, the bombers will momentarily drop off the radar screen. Since the range is very close (10-12 nm), it requires the crew to accurately "predict" the actions of the bombers and correctly use their radar so that they may be quickly re-acquired. The problem is further complicated in that the bombers and fighters will now "merge" in roughly the same airspace. If the fighters are ignored, then they can launch against the F-15s. If the F-15s "lock" their radar on the fighters, which will usually be the case at this point, then the bombers can continue toward the airfield "untargeted." Once the fighters are engaged, it is very difficult to re-acquire the bombers since they are low and will be flying away from the F-15s. If the F-15s fail to kill the fighters, the problem will only be compounded.

This example not only shows the approach taken toward the design of the mission scenarios, but also serves to illustrate our contention that SA is more than knowledge of the current situation. In operational environments, situation assessment and decision making are viewed as tightly coupled and are often difficult to separate. For the fighter pilot to be successful, he must not only be able to "build the big picture," but he must also translate his assessment into an employment decision. Often, the inability to make these critical employment decisions may lead to mission failure despite a correct assessment of the situation. In the sample scenario, the key to success is to target and destroy the bombers prior to 15 nm and then target the fighters. If the ranges become so close that all four threats must be dealt with simultaneously then the mission is likely to fail. It is through the careful design of such mission scenarios that the failure to incorrectly assess the situation or make incorrect employment decisions can be

successfully inferred based upon the observation of pilot performance in the unfolding of the mission scenario.

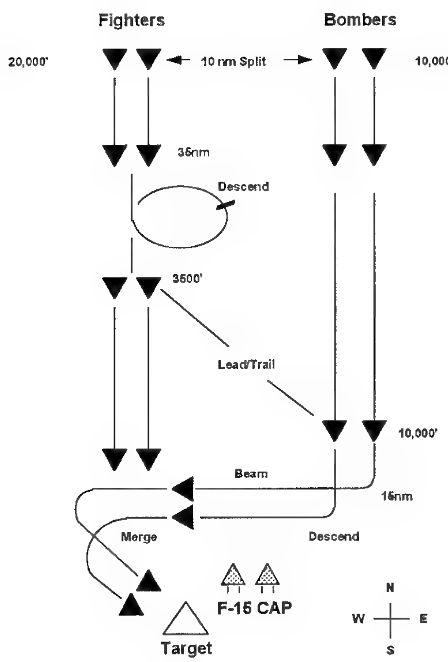


Figure 2. Typical Engagement Scenario

Data Sources. Given the tremendous cost of gathering data on MR F-15 pilots, the approach was to gather as much as possible from a variety of sources. In our view, the most important data sources were the judgments and observations of two retired fighter pilots who possessed an in-depth understanding of the air combat domain. The same two SMEs were used throughout the year-long data collection effort. For each mission, the following procedure was followed. One of the SMEs would attend the mission briefing session conducted by the crew. During the conduct of each mission, both SMEs observed mission performance. One of the SMEs also served as the mission director who was responsible for starting and stopping each engagement, communicating with the console operator, etc. During each engagement, each SME independently completed an observational checksheet to record pertinent events, notes, and outcomes. Upon completion of the four engagements comprising a single mission, one of the SMEs accompanied the crew to the debriefing room. The flight lead was responsible for conduct of the debriefing, although the SME was permitted to ask questions in an attempt to clarify the crew's understanding of the situation and purpose of their actions. Upon completion of the debrief, the two SMEs discussed each engagement, and completed a consensus performance rating scale consisting of the 24 behavioral indicators of SA related to F-15 mission performance. The SMEs also produced a written critical events analysis for each mission which attempted to identify those events that, in their opinion, affected the outcome of the mission and were indicative of the crew's SA.

A variety of other data were also gathered. These included mission events and outcomes such as weapons firings, kills, etc. Using the data logger in the ECS, the digital data passed over the network was recorded, whereby each engagement could be reconstructed. The videos recorded and used for debriefing were also archived. Additionally, eye movement data were recorded for the four engagements flown on the last mission. And finally, all participants were also asked to "critique" the simulation and also give opinions regarding its potential for training.

3 RESULTS

The results from two data sources are presented in this paper, the performance ratings from the two SMEs, and the critiques regarding the potential value of the simulation for training. These data are used to address the two issues central to this paper, namely, the use of simulation as a tool for both measuring and training SA.

Simulation as an SA Measurement Tool. One of the original goals of the overall research program was to develop techniques for measuring SA. In essence, two approaches were taken; first, the development of SA rating scales that could be administered within the operational units; and second, the development of techniques based upon observed performance within a controlled simulation environment. To briefly summarize the first approach (2), three SA Rating Scales (SARS) were developed to measure pilot performance in an operational fighter environment. These instruments rated SA from three perspectives: supervisors, peers, and self-report. SARS data were gathered from 239 mission-ready USAF F-15C pilots from 11 operational squadrons. Reliabilities of the SARS were quite high as measured by their internal consistency (.97 to .99) and inter-rater agreement (.84). Correlations between the supervisory and peer SARS were strongly positive (.85 to .87), while correlations with the self-report SARS were positive, but smaller (.50 to .58). A composite SA score was developed from the supervisory and peer SARS using a principal components analysis. The resulting score was found to be highly related to previous flight experience and current flight qualification. In fact, this score was used as the basis for selection of pilots to participate in the simulation phase of the effort that is described here.

One question of interest is the relationship between the SA scores based upon peer and supervisor ratings in the squadron and the SA scores derived from the simulation environment. The hypothesis was that there would be a moderately positive correlation between these two sets of scores. Simply stated, pilots judged to perform very well in the units should also perform well within a controlled simulation environment and vice versa. Mean performance ratings given by the SMEs across the four engagements were computed for each mission. These mission ratings were then regressed against the single

score obtained from the units. A scatterplot of these data are presented in Figure 3. The resulting correlation was found to be .56 (p<.01).

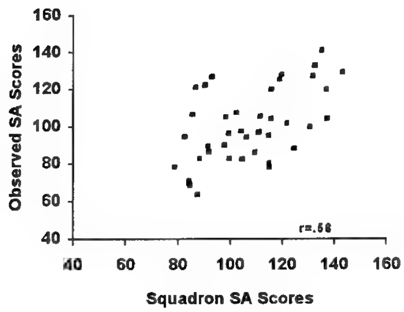


Figure 3. Scatterplot of SA Scores in Squadron Versus SA Scores in Simulator

These results support the hypothesis of a relationship between SA as measured in both environments. They also indicate that although the relationship is positive, it is not perfect.

Simulation as an SA Training Tool. The other issue concerned the potential of multiship simulation as a tool for training SA skills. Given the definition of SA that was adopted at the outset of the study, this question translates into the issue of whether training in this type of simulated combat environment transfers to the real airborne environment. While transfer is an easy concept to understand, it is extremely difficult to measure given the enormous costs and complexities of carrying out such evaluations.

Bell and Waag (12) have proposed a five-stage sequential evaluation model for conducting training effectiveness evaluations. In order, these include: (1) utility evaluation; (2) in-simulator performance improvements; (3) transfer to alternative simulation environment; (4) transfer to a flight environment; and (5) extrapolation to a combat environment. The authors made use of a multiship combat simulation similar to that used in this study as a vehicle for discussion of the requirements of each of these stages. The data gathered from the study presented bear only upon the first two--user opinion and performance improvement.

Two types of user opinion data were gathered--ratings of the training benefit for various pilot experience levels and an open-ended questionnaire. The results of the ratings of potential training benefits are provided in Figure 4. These data clearly indicate that positive opinions were expressed by the study participants on the value of this type of simulation for training. The potential training was considered beneficial for all levels of qualification. However, as expected, greater benefit would be expected for pilots upgrading into a given qualification level.

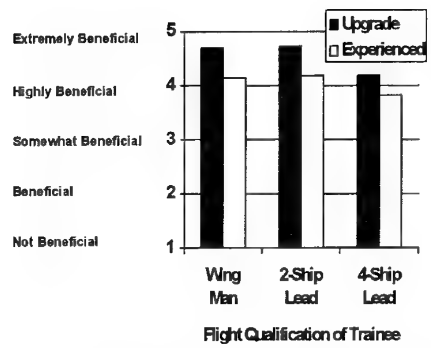


Figure 4. Rated Benefit of Training for Various Levels of Experience

Opinions expressed in the open-ended questionnaire were also quite positive. Although qualitative, they provide additional insight into the potential focus of training using multiship simulation and how it might be employed. In particular, mention was made of using such training as a means of enhancing both situation assessment and decision-making skills. It was also frequently noted that there was tremendous value in learning flight leadership and resource management skills. In terms of the location of such simulation, the overwhelming consensus was that they would be of most value within the operational units. This was not too surprising since each unit now has the operational version of the cockpits used in the present investigation. However, they are stand-alone and non-visual, and as such their training capability is fairly limited. In contrast, the networking of such devices within a realistic combat environment increases the potential greatly. The bottom line from the utility data is that the participants considered multiship simulation as a tool with high training potential.

While positive user opinion is a necessary prerequisite for effective training, in itself, it is insufficient validation (12). At the next stage of the evaluation model, it is necessary to demonstrate improved performance within the simulation environment as a function of practice. In other words, it is necessary to show that learning has occurred. It should be pointed out that it was never the intent, at the outset of the study, to demonstrate performance improvements. It must be emphasized that the sole purpose was to develop a set of simulation scenarios that could be used to assess SA within a combat environment. As such, normal training interventions were not permitted. For example, during the debrief, pilots were permitted to only view their own in-cockpit displays and not the planned view display. Moreover, the two SMEs were not permitted to provide any type of feedback to the pilots regarding their performance.

However, data from the ninth mission did permit some comparison since identical scenarios had been flown earlier in the week. The ninth mission was designated the "eye track" mission in which eye movement data was recorded. For these

scenarios, an eye tracker computed point of gaze and was displayed against the background scene as determined from a scene camera mounted on the pilot's helmet. The resulting video signal replaced the second cockpit display within the ECS. This permitted the crews to debrief the final mission using three integrated displays, the planned view of the fight, their own cockpit display, and the eye-tracked display which portrayed point of gaze against the background scene. Although not central to this paper, it should be mentioned that very positive opinions were expressed by the pilots regarding the potential of eye movement recordings as a feedback tool for training. It was viewed as potentially useful for the earlier stages of training and, in particular, for the diagnosis of problems of students encountering difficulty. It could potentially provide a solution to the continuing problem of training for single-seat aircraft in which instructors complain that diagnosis is difficult when one cannot see where the student is looking.

Two scenarios, a 2 V 2 defensive counter air (DCA) mission and a 2 v 4 offensive counter air (OCA) mission, were flown during the middle of the week and then again on the last mission. A comparison of performance is presented in Figure 5. In both cases, performance on the last mission was improved. However, only the 2 V 2 DCA mission was found to be statistically significant.

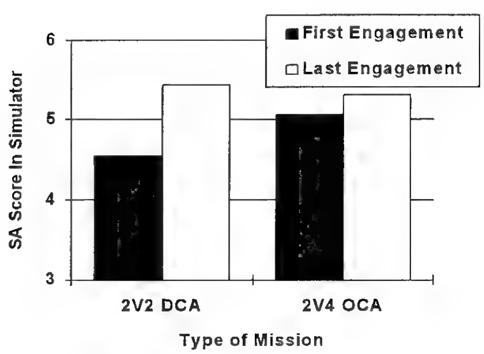


Figure 5. Effects of Practice on Observer SA Ratings

It should be recalled that the scenarios were designed to increase in difficulty over the week. Consequently, if one simply plots the Observer SA Scores across missions, there is generally a downward trend. To obtain an estimate of what the curve might look like assuming "equal difficulty" of all scenarios, a magnitude estimation procedure was undertaken to scale the difficulty of the scenarios. Raters included the two SMEs and another in-house F-15 pilot who had occasionally served as wingman in the course of the study. Only missions 2 through 8 were included since mission 1 was a "familiarization" sortie and mission 9 was the eye track sortie. These difficulty weightings were then applied to the mean observed SA scores for each mission. The results are presented in Figure 6.

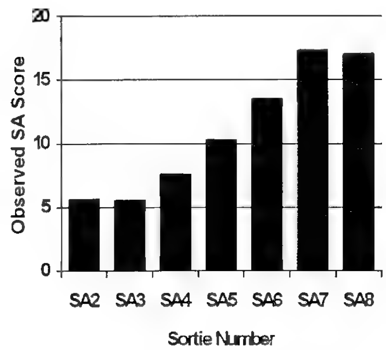


Figure 6. SA Scores Weighted for Scenario Difficulty Across Missions

It is clear that when the scores are weighted for scenario difficulty, the resulting curve suggests that performance improved over the week. Again, it should be cautioned that the procedures followed were not the most appropriate for a conduct of a rigorous test of learning within the simulation environment. However, when such data are coupled with the very strong pilot opinions that they had received valuable training, it seems reasonably safe to conclude that learning had occurred over the week.

4 DISCUSSION

This study attempted to answer two questions. Can multiship simulation be used as a tool for both measuring and training SA? Each of these is discussed. The reader should keep in mind the operational definition of SA that was adopted at the outset of the investigation since it does markedly differ from others that have been used.

First, can multiship simulation be used as an assessment tool? In our view, the answer is clearly "yes." The data presented in this paper show a positive relationship between SA as measured within the operational units and SA as measured in a controlled simulation environment. Although the relationship is positive, it is not perfect. The data from the units were found to relate very strongly to previous flight experience and current flight qualification. In general, the same relationships were observed in the simulation data, although their magnitudes were reduced. Those pilots with more flight hours and a higher flight qualification, in this case an instructor pilot rather than a 2-ship flight lead, generally performed better. As a group, the best performers were those pilots who were weapons officers, indicating that they were Fighter Weapons School graduates. Taken as a whole, these data suggest that those pilots with more experience tend to perform better within a controlled simulation environment.

However, there occurred noticeable exceptions to this general trend. For example, consider the three pilots in Figure 3 who had low squadron scores but performed extremely well in the simulation environment. These individuals were fairly

inexperienced two-ship leads and for that reason obtained low squadron scores. However, these individuals adapted extremely well to the demands of the scenarios that were used in the study. In other words, they learned very quickly and adapted to the demands of the combat environment. In fact, their performance was superior to other pilots who were certainly more experienced. It should be emphasized that the scenarios flown on the last four missions were of a complexity that is rarely experienced within operational training environments due to resource constraints. Although speculative, such data suggest that simulation may be a useful tool in assessing not only current performance, but also predicting who is likely to excel in new environments for which they have not received training.

Second, can multiship simulation be used as a training tool? In our view, the answer is, again clearly "yes." From a user's perspective, the data are very clear regarding the potential value of such simulation for training. The 63 MR F-15 pilots overwhelmingly considered such training to be of value. Although such anecdotal evidence is often considered suspect from a scientific perspective, it is nevertheless an absolute prerequisite for effective training. Unless there is user acceptance, the resulting training will be of marginal value regardless of the device's inherent potential.

In addition to the opinion data, there is evidence that performance did improve within the simulation environment; in other words, learning did occur. Again, it should be pointed out that the amount of improvement was probably "minimized" due to the evaluative orientation of the investigation. When identical scenarios were flown early and late during the week, the performance on the second repetition was better. Additionally, when scenario difficulty is assumed constant, the resulting weighted scores show improvements. These data combined with the fact that the study participants expressed opinions to the effect that their proficiency had improved leave little doubt that learning had occurred.

Although the data clearly indicate (a) that the end user expresses very positive opinions toward the value of multiship simulation and (b) that learning occurs, there still remains the issue of transfer to the real world which represents the "acid test." Clearly, the data gathered in this study do not bear upon that issue. For the "believer," evidence to date is strong enough to warrant the conclusion that training will be effective. In fact, given the previous transfer of training research that has already been conducted (12,13), there is little reason to suspect that such training within a multiship simulation environment would not have a positive effect upon subsequent performance in the air. Yet, for the "skeptic," no definitive evidence has been presented.

Based upon the findings of the present study, it is concluded that multiship simulation can be successfully used as a tool for both measuring and training SA. Future efforts should focus upon the development of appropriate training strategies and interventions which will maximize its training potential.

5 REFERENCES

1. McMillan, G.R., Bushman, J. & Judge, C.L.A. (1995) Evaluating pilot situational awareness in an operational environment. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

2. Waag, W.L. & Houck, M.R. (1995) Development of criterion measures of situation awareness for use in operational fighter squadrons. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

3. Carretta, T.R. & Ree, M.J. (1995) Determinants of situational awareness in U.S. Air Force F-15 pilots. In AGARD-CP-575, Situation Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

4. Carroll, L.A. (1992). Desperately seeking SA. TAC Attack (TAC SP 127-1) 32: 5-6.

5. Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Monica, CA: Human Factors Society.

6. Fracker, M.L. (1988). A theory of situation assessment: implications for measuring situation awareness. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 102-106). Santa Monica, CA: Human Factors Society.

7. Houck, M.R., Whittaker, L.A., & Kendall, R.R. (1993). An information processing classification of beyond-visual-range intercepts. (AL/HR-TR-1993-0061). Williams AFB, AZ:Armstrong Laboratory.

8. Gehl, T.L., Rogers, R.L., Miller, M.A., & Rakolta, J. (1993). Interdependence of training utility and network performance using the Armstrong Laboratory Multiship Research and Development System. In, Proceedings of 15th Industry/Interservice Training Systems Conference, Orlando, FL: National Security Industrial Association.

9. Platt, P. & Crane, P. (1993). Development, test, and evaluation of a multiship simulation system for air combat training. In, Proceedings of 15th Industry/Interservice Training Systems Conference, Orlando, FL: National Security Industrial Association.

10. Rogers, B. (1992). Tactical air threat system for a distributed simulation network. In, Proceedings of 14th

Industry/Interservice Training Systems Conference, San Antonio, TX: National Security Industrial Association.

11. Tenney, Y.J., Adams, J.J., Pew, R.W., Huggins, A.W.F, & Rogers, W.H. (1992). A principled approach to the measurement of situation awareness in commercial aviation (NASA Contractor Report 4451). Langley VA: National Aeronautics and Space Administration.

12. Bell, H.H. & Waag, W.L. (1994). Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A Review. Submitted to International Journal of Aviation Psychology.

13. Waag, W.L. (1981). Training effectiveness of visual and motion simulation. (AFHRL-TR-79-72).Williams AFB, AZ: Air Force Human Resources Laboratory.

Situation Awareness and Workload: Birds of a Feather?

K.C. Hendy
Defence and Civil Institute of Environmental Medicine
PO Box 2000
North York, Ontario M3M 3B9
CANADA

1. SUMMARY

In this paper it is argued that an hierarchical information processing model, with a basis in perceptual control theory, provides the necessary framework for interpreting a large, unfocused empirical literature on the topics of workload and situation(al) awareness (SA). The fundamental importance of situation awareness will emerge in considering the role of the mental model in providing the reference signal for a closed loop perceptual control system. It will be asserted that those aspects of the mental model generally covered by the SA rubric result from high level information processing activity that requires spare capacity to service. Increasing time pressure (workload) reduces the capacity available for this activity. An experiment in the application of a workload scale (NASA TLX) and a situation awareness metric (SART) to a simulated air traffic control environment is cited. It will be shown that the situation awareness scale taps largely into the workload side of the equation rather than the SA side. Implications for the measurement of SA will be drawn.

2. INTRODUCTION

Consider the following statement

In the ideal cockpit we would like aircrew to develop high levels of situation awareness using their cognitively compatible displays while experiencing low levels of task induced workload and achieving optimal, error free, performance.

Is this a reasonable goal for the systems designer? What are the relationships between the hypothetical constructs of *situation awareness* and *workload* and how is *performance* dependent on these concepts? Is *cognitive compatibility* part of this puzzle, and where is the theoretical framework that binds these ideas together? This paper attempts to provide such a framework, and argues that these concepts are less *birds of a feather*, but rather they are components of the same bird.

In many industrial and military systems, the potential for an operator to perform effectively when responding to novel situations such as malfunctions, emergencies, and unexpected occurrences depends on their knowledge of the moment to moment changes in the status of pertinent system variables, their deviation from a set of desired states or goals, the dynamics of the controlled system and the interactions between system variables. This knowledge forms an *internal representation* or *mental model* of the process to be controlled. The concept of a *mental model*, which the operator develops and draws upon when making operational decisions, is central to the idea of *situation awareness*, and has become an aspect of particular concern to engineers and behavioural scientists involved in the development of complex human-machine systems.

While measures of performance and workload have been the typical metrics employed for determining the efficacy of human-machine interactions, there are certain conditions under which these measures are limited (see, for example, the work of Yeh and Wickens [1]). Take, for example, a situation in which the optimum strategy for an operator is to simply wait and monitor system variables before deciding whether or not to take action. In this situation there may be no overt performance to measure but cognitive load may be high. Further, consider a situation where an operator is flooded with activities, or the converse, where workload is relatively low and the operator is performing a passive monitoring role. Each of these scenarios, though arguably opposite in terms of their levels of workload, may produce a state of low situation awareness. In the former case the operator may have little spare capacity to develop a mental model while in the latter case the operator may be *out-of-the-loop* and lacking both relevant information, and a *feel* for the system dynamics which are essential to building the knowledge state that would allow an effective intervention. Because of the potential difficulty in determining operator effectiveness under these types of conditions, one might speculate that the concept of the mental model may help provide relevant information about an operator's *potential* to perform effectively in certain types of complex systems.

Therefore while workload, and situation awareness appear both to be relevant to human performance, their synthesis through theory has been sadly lacking. This paper outlines an attempt to build an integrating framework for *workload*, *situation awareness* and *performance* from two theoretical models, namely, Hendy, Liao and Milgram's [2] Information Processing (IP) Model and William T. Power's Perceptual Control

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

Theory or PCT [3]. A new construct, termed *cognitive compatibility*, will be interpreted within this framework. Brief mention will also be made of empirical investigations that have looked at the relationship between *workload* and *situation awareness* as measured by the NASA Task Load Index or TLX [4], and the Situational Awareness Rating Technique or SART [5].

3. WORKLOAD

3.1 The IP Model

In Miller's words (reprinted as [6]), "...Insofar as living organisms perform the functions of a communication system, they must obey the laws that govern all such systems..." Using an information processing paradigm, the IP Model attempts to provide a coherent theory for synthesizing much of the literature on workload and performance. The dependency of workload, performance and errors on *rate of processing*, is central to this model. For a more complete description of the IP Model, and the predictions that flow from it, see [2, 7, 8].

It can be shown from the IP Model, if the operator adopts a constant problem solving strategy, that workload and performance are both driven by the ratio:

$$\frac{\text{time taken to process the information necessary to make a decision}}{\text{time available before the decision has to be actioned}}$$

This ratio provides a measure of the *time pressure*. The IP Model posits that performance, errors and subjective experiences of workload are all determined by *time pressure*.

The IP Model is a dynamic model, which predicts that an operator will adapt to excessive time load by two fundamental mechanisms, namely: (1) by reducing the amount of information to be processed; or (2) by increasing the time before the decision must be actioned. These mechanisms are attributed to changes in processing strategy, with such adaptations usually involving a trade-off between the amount of information processed and the achievement of an acceptable level of performance. Any particular problem solving strategy is assumed to involve certain processing structures at the neural level, with multiple *concurrent* tasks competing first for specific processing structures, and then for time [8]. A given structure is assumed to process in a time multiplexed serial fashion. It is assumed that the actual processing rate within a structure remains more or less constant [9], although the possibility that processing rate is affected by changing physiological states, brought on say by fatigue, is allowed.

While workload is generally regarded as multi-faceted, the IP Model reduces the effects of all factors that contribute to cognitive load either to their influence on the *amount of information to be processed* or to their effect on the *time allowable* before a decision has to be implemented.

3.2 The Relationship Between Workload and Performance

The IP model explicitly associates degraded performance either with the information directly shed if adaptation does not bring the time pressure below 1 or, alternatively, with the selection of a strategy that results in more rapid but less precise action (both situations involve information, which is relevant to the performance of the task, left unprocessed). Hence, performance and errors are inextricably and predictably tied to the imposed time pressure.

In the IP Model it is also assumed that operators respond to some function of time pressure when reporting subjective experiences of workload. With this assumption, a relationship between performance (defined specifically in the IP Model as the ratio between the *amount of information processed* to the *amount necessary for error free performance*) and operator workload is established through their common dependency on time pressure.

4. SITUATION AWARENESS

4.1 A Working Definition

In any activity, information is processed within the structure of the situation that the operator is immersed in. Knowledge of this situation gives context to the decisions that are made and gives form to the actions that are taken. In turn, this determines the appropriateness of the responses. Knowledge is resolved uncertainty. Hence, knowledge reduces the amount of information that must be processed in arriving at a future decision. This is the realm of Situation Awareness (SA). For the purposes of discussion, consider the following definitions:

The Mental Model is that part of the operator's internal state which contains the knowledge and structure necessary to perform a task. As such, the operator's mental model directly shapes the operator's actions and determines the potential to perform in accordance with the system demands. The mental model contains the operator's goal state and provides the reference against which actions are selected and initiated.

The term Situation Awareness (SA) particularly relates to that dynamic and transient state of the mental model which is produced by an ongoing process of information gathering and interpretation during the performance of some job of work. While the concept can be generalized to all tasks, no matter what their complexity, the term SA is usually used when considering tasks that have strategic and tactical components such as flying an aircraft, controlling or monitoring a plant, or tactical decision making.

These definitions emphasize the role that the mental model plays in shaping perception and action in goal-directed human activity.

4.2 Perceptual Control Theory

The role of feedback in goal-directed human activity, is a fundamental tenet of William T. Power's Perceptual Control Theory [3]. Powers' model is organized hierarchically with many goals providing the reference points for multiple layers of control; from the lowest levels of processing up to abstract goals such as the need for self esteem and actualization. In the PCT model, an action or behaviour is emitted in response to an error correcting signal that is transmitted with the intention of changing the state of the world so that the operator's perception matches a desired state or goal. The fundamental claim of PCT is that it is the perception that is controlled, **not** the behaviour. As behaviour is not the controlled quantity, one should expect considerable variability between and within individuals.

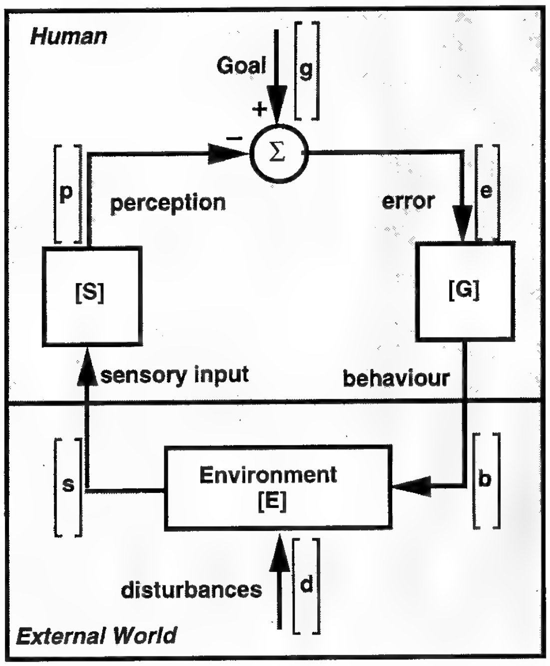


Figure 1. *William T. Power's Perceptual Control Model.*

Power's PCT model is represented diagramatically in Figure 1. The hierarchy of control is represented using a matrix formulation. The hierarchy of *goals*, *errors*, *behaviours*, *disturbances*, *sensory inputs* and *perceptions* are shown in vector form in Figure 1 (i.e., **g**, **e**, **b**, **d**, **s**, **p**), while the transfer functions **G**, **E** and **S** are shown as matrices. In general, **S** and **G** will have latencies or transport delays associated with the requirement to process information. These latencies have already been described in terms of the *decision time* in the IP Model. Transport delays effectively add an additional lag term to the loop which slows the rate at which the loop can respond to null an error state. The dynamics of the external world are contained in **E** (the characteristics of the vehicle or plant, the tactics of the opposing forces, the user interface, etc.).

From Figure 1, it can be seen that perceptions and actions are shaped by the transfer functions **S** and **G** as follows

$$\begin{bmatrix} e \end{bmatrix} = \begin{bmatrix} g \end{bmatrix} - \begin{bmatrix} p \end{bmatrix}$$

$$\begin{bmatrix} p \end{bmatrix} = \begin{bmatrix} \quad S \quad \end{bmatrix} \begin{bmatrix} s \end{bmatrix}$$

$$\begin{bmatrix} b \end{bmatrix} = \begin{bmatrix} \quad G \quad \end{bmatrix} \begin{bmatrix} e \end{bmatrix}, \text{ and}$$

$$\begin{bmatrix} s \end{bmatrix} = \begin{bmatrix} \quad E_1 \quad \end{bmatrix} \begin{bmatrix} b \end{bmatrix} + \begin{bmatrix} \quad E_2 \quad \end{bmatrix} \begin{bmatrix} d \end{bmatrix}.$$

One can associate the goal state **g** and the transfer functions **S** and **G** with the operator's mental model. In fact if the set of all $\mathbf{g} = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n\}$ represents all possible goal states, the combination of **S**, **G** and $\{\mathbf{g}_i\}$ could be considered to *be* the operator's mental model. It is expected that **S**, **G** and **g** will not be static but will change with time as learning and adaptation take place. The transfer matrices **S** and **G** contain all the transformation rules and relationships (the knowledge) that allows one to operate on the environment **E** in such a way that the perceived state of the external world can eventually be made to match the internal goal state. As the degrees of freedom for sensory input will be much greater than the degrees of freedom of the emitted behaviours, **S**, **G** and **E** will not be square.

4.3 The Relationship Between Situation Awareness, Performance and Workload

This interpretation of the mental model, in terms of a vector of goal states **g** and the transfer functions **S** and **G** of a multi-layered perceptual control loop, quite clearly illustrates the central role the mental model has in shaping both perception and action. The mental model contains stable long term memory relationships but also changes dynamically as the loop adapts to the transient aspects of the current situation. Note that this adaptation will only apply to those variables that are being actively controlled or attended to (the concept of *active control* does not require an overt action to be emitted as internal *imagination* loops are postulated). Hence, SA is gained over time through interaction with the environment (either *real* or *imagined*). Applying the IP Model to the transformation matrices **S** and **G**, one would argue that the transport delays experienced,

in forming percepts from sensory inputs and in emitting actions from error states, will depend on the amount of information that has to be processed in going from **s** to **p** and from **e** to **b**.

Efficient and rapid processing implies appropriate strategies that involve small amounts of information to be processed (i.e., prior knowledge is used to reduce the uncertainty of the current situation, through the use of skill-based behaviours [10]; or Klein's recognition-primed decision making [11]). These strategies come from higher order knowledge, such as the relationships between things, and the integration of individual items into patterns. In a changing environment, the development of this knowledge is a task that demands attentional resources to service. Hence, SA and workload are obviously related to the extent that the development of these aspects of the mental model will depend on the availability of processing resources for the active control of these higher order loops.

In periods of overload, spare capacity may not be available to service these high level loops. Therefore while a high level of SA has the potential to reduce the amount of processing associated with some future decision, and hence reduce time pressure, it consumes processing capacity in the period leading up to that decision. When the workload comes from the control of loops that do not involve the variables associated with higher order SA, high workload will detract from the development of SA. Alternatively, if the workload involves the control of loops that involve the SA variables, high levels of workload may be associated with a well developed mental model. Hence, workload and SA are likely to dissociate.

4.4 Ramifications for Measurement

The definitions offered for SA in this paper suggest that an appropriate experimental paradigm for measurement would involve forcing a subject to make a decision, through some intervention, which is based on an understanding of the current state of some dynamic situation. This decision should be at the level of rule- or knowledge-based behaviour to be of interest. The key to this paradigm is the forcing of an action (performance) in order to test the operator's internal representation.

The manifestation of SA will be seen in the timeliness and appropriateness of the subject's decision(s) following the intervention (failure of an automatic system, retasking etc.). The word *appropriateness* rather than *correctness* is used here because a variety of actions can cause the error signal eventually to be nulled. All that is required, for effective and complete error correction, is that the loop gain be negative and >> 1. Other measurement techniques might include verbal protocols, or probes directed at eliciting the knowledge (the mental model) which is considered important to decision making (e.g., through the Situational Awareness General Assessment Technique — or SAGAT [12] — or similar methods).

Note that the timeliness of goal achievement depends both on the strategy used (as determined by the transformation terms selected from the transfer matrices **S** and **G**) and on the phase characteristics of the loop gain **SEG**. Actions that are *appropriate* will result in a high correlation (in the sense of zero phase error) between **p** and **g**. It is the role of training to develop an appropriate repertoire of primed perceptions **s→p** and actions **e→b**. Therefore, while Powers suggests that the observation of behaviour is not a good indicator of goal-directed human activity [13], it seems that a range of normative and, in the sense discussed above, appropriate behaviours can be defined for many situations. Obviously this requires that goals have been clearly and unambiguously established.

5. COGNITIVE COMPATIBILITY

5.1 A Definition

Far less mature than the concepts of workload and situation awareness, the hypothetical construct of *cognitive compatibility* has been coined recently. Consider the definition [14]:

[The] Cognitive compatibility of advanced aircraft displays is the facilitation of goal achievement through the display of information in a manner which is consistent with internal mental processes and knowledge, in the widest sense, including sensation, perception, thinking, conceiving and reasoning.

5.2 The Relationship Between Cognitive Compatibility, Situation Awareness and Workload

The cognitive compatibility of a display can be interpreted in terms of the match between the characteristics of the display as represented by the sensory vector **s** and that part of the operator's mental model, contained in the matrix **S**, which operates on this sensory input. A cognitively compatible display would invoke only terms of **S** that result in the highest gain×bandwidth product possible. Thus, the *cognitive compatibility* of a display will be manifested in the time taken for goal achievement from the onset of some sensory input. From the IP Model, this translates directly into the timeliness and appropriateness of the emitted action(s).

This forges the link between cognitive compatibility and both workload (through the frequency domain) and the mental model (through **S**). Note that in observing behaviours, the effects of **g**, **G** and **E** are confounded with the effects of **S**. Hence, appropriate controls must be exercised in trying to separate the effects of cognitive compatibility from effects of changes in goals, strategy/response selection, or the external environment.

6. AN EMPIRICAL STUDY

An experiment was run to investigate the relationship between operator workload and situation awareness as measured by the NASA TLX and SART respectively. Of course such an experiment does not necessarily test

the relationship between operator workload and SA, but merely investigates the relationship between two measurement instruments that are intended to capture aspects of these concepts.

6.1 The Task

The experimental task was a simulated Air Traffic Control environment. The task, called ATC 2.0, was an early version of a computer game which is available from the internet and various bulletin board services. Briefly ATC runs on a Macintosh computer and presents a simulated radar screen on which aircraft targets and the locations of airports are shown. The numbers of aircraft, airports and the session time are set by the experimenter. Aircraft arrive and depart at the 8 cardinal points of the compass as well as at airports. Flight paths (headings and altitudes) are controlled with a mouse using soft keys on the screen.

6.3 Results and Discussion

The individual scale data from the TLX (6 scales) and the SART (10 scales) was subject to principal component analysis using SYSTAT version 5.2 for the Macintosh [16]. The resulting unrotated factor loadings are shown in TABLE 1. Factor loadings less than 0.5 are omitted for clarity. The first three factors together explain 69% of the variance. Varimax rotation spread the variance over more components but did not appear to yield a more interpretable structure.

The 16 scales in TABLE 1 were categorized according to their contribution to *Resource Demand*, *Resource Supply* or *Understanding* using the same taxonomy that Selcon and Taylor [15] used for SART. Lacking a theoretical rationale, this categorization is rather arbitrary. While the *Resource Demand* factors have some degree of face validity, the *Resource Supply* factors are more difficult to rationalize.

TABLE 1:

Unrotated factor loadings from the principal component analysis of the pooled TLX and SART scale data (factor loadings < 0.500 are omitted). The first three principal components (PC1, PC2, and PC3) are shown.

Scale	Origin	PC1	PC2	PC3
Resource Demand				
Mental Demand	TLX	0.917		
Physical Demand	TLX	0.517		-0.590
Temporal Demand	TLX	0.892		
Effort	TLX	0.912		
Instability	SART	0.662		
Complexity	SART	0.847		
Variability	SART	0.920		
Resource Supply				
Frustration	TLX	0.545	-0.577	
Performance	TLX	0.569	-0.534	
Arousal	SART		0.582	
Concentration	SART	0.857		
Division of Attention	SART		0.627	
Spare Capacity	SART	-0.765		
Understanding				
Quantity of Information	SART		0.738	
Quality of Information	SART		0.801	
Familiarity	SART			0.571

6.2 Subjects and Method

Ten subjects participated in the experiment. Sessions lasted 15 minutes. Twelve schedules were created with the number of aircraft arrivals ranging from 5 to 25. Arrivals at the eight cardinal points, and departures from airports, occurred randomly during the session time. At the termination of the 15 minute session the NASA TLX and the 10 dimensional SART [15] were administered.

In many cases the distinction between a supply factor and a demand factor is ambiguous. Lacking a definition of a *resource* it is difficult to say what factors might result in their greater availablity.

In TABLE 1, the *Resource Supply* category is a mixture of emotional, global activating, and attentional factors. It is not clear for example whether subjects, in rating the scales, would see *Concentration, Division of*

Attention and *Spare Capacity* as driven directly by the task demands. If this were the case then this would place them on the *Demand* side rather than the *Supply* side of this taxonomy.

In terms of the IP Model, the resource that is being managed is time. Factors, such as *frustration, fatigue, mood, knowledge of one's own performance, arousal, motivation* etc. are claimed, in this model, to modulate the subject's efforts in adapting to increasing time pressure through the use of more time efficient strategies. From the IP Model, the role of attentional factors such as *Concentration* and *Division of Attention* in determining the supply of processing resources, is likely to be indirect.

It can be seen from TABLE 1 that the first principal component appears to be a demand factor. Although the *Spare Capacity* scale was originally categorized in the *Resource Supply* class, its loading on PC1 suggests that subjects were rating this scale in terms of (1 - *Demand*). Hence, this scale is perhaps more correctly thought of in terms of *Resource Demand* rather than *Resource Supply*. Similarly, subjects may have interpreted the requirement to concentrate as a manifestation of the task demands.

Factors associated with the *Quality* and *Quantity of Information* load most heavily on PC2. With *Concentration* and *Spare Capacity* shifted to the *Resource Demand* side, the remaining *Resource Supply* factors load partially along the directions of both PC1 and PC2. Therefore, in summary, two main factors emerge: (1) a demand or workload-related factor; and (2) a factor largely related to acquired knowledge (this could be termed the SA factor). It should be noted that the manipulation used in this experiment, and in the other experiments refered to in this paper, was mainly a workload manipulation. Not all factors of the TLX and SART scales were manipulated, either directly or indirectly, to create the variances necessary to fully identify the underlying structure of these instruments.

Overall the pattern of results from the ATC experiment is similar to that found by Selcon and Taylor [15]. One interpretation that may be offered for these results is that with the exception of the *Quality* and *Quantity of Information* scales (and possibly also the *Familiarity* scale) SART is largely a workload instrument. In the words of Selcon, Taylor and Koritsas [17] "*...It can be concluded...that both the TLX and SART are sensitive to changes in task demands, and that they appear, along this dimension, to measure the same things.*" They go further to draw the following conclusions "*...This could be taken as evidence that there is commonality, not just between the scales, but also between the concepts of workload and situational awareness.*" While the conclusion that SART and TLX instruments may measure much the same thing seems defensible, extrapolating to equate the concept of SA with workload does not appear to be justified. For this argument to be sustained it would have to be proved that TLX and SART are truly measuring what they purport to be, namely workload and SA respectively.

From the IP and PCT models, SA and workload can be seen as two independent aspects of human information processing. This theoretical position might be seen reflected in the pattern of weights from the first two principal components obtained both in the ATC experiment and in Selcon and Taylor's 1989 experiment. Yet despite this underlying independence, workload and SA are totally bound together albeit in a potentially predictable fashion.

6. DISCUSSION AND CONCLUSIONS

The combination of the IP Model and Perceptual Control Theory provides a coherent framework for tracing the relationships between concepts such as workload, situation awareness and cognitive compatibility. From this theoretical position one can talk about workload in terms of a readily understandable and measurable quantity termed *time pressure*.

Also emerging from this approach is the dominance of the mental model in shaping all goal-directed human activity. Rather than being a facilitator of action, there can be **no** action without the involvement of the mental model. Combining the IP Model with PCT, the relationship between workload and SA can be seen manifested in transport delays as sensation maps into perception and perceived error states are mapped into action. On the other side of this equation is the requirement for attentional capacity to be available so SA can be learnt in dynamic situations. Building the dynamic and transient knowledge associated with SA requires active control of the high level processing loops that use this knowledge for forming perceptions from sensory inputs and for shaping actions in response to perceived error states. In order to assess the state of this knowledge, these transformation rules and relationships must be made to operate, either by forcing an overt action or by knowledge elicitation techniques. Good SA is associated with rapid goal achievement through timely and appropriate actions in response to some sensory input. The mental model, in general, represents the organism's adaptation to the environment.

Cognitive compatibility is traced to the match between the sensory vector and the transformation relationships that form perceptions from this input. A high level of cognitive compatibility would facilitate goal achievement through timely and appropriately formed perceptions. Cognitive Compatibility is a property of the interface between the human and the environment, and represents an attempt to adapt the environment to be consistent with those terms of the organism's mental model that result in timely and appropriate actions. Therefore, cognitive compatibility has aspects of both consistency with the mental model and outright performance (in terms of a high gain×bandwidth product) associated with it. Both aspects must be

satisfied for a display to be accepted as cognitively compatible.

Finally, because of the fundamentally separate and distinct nature of workload and situation awareness these two concepts should be treated and measured separately. However, because both workload and SA combine in their effects on task performance, attempting to validate metrics that are composites of workload and SA factors against performance is difficult. While it is workload, through time pressure, that ultimately determines performance and error rate according to the IP Model, the time domain behaviour of the perceptual control loops is entirely bound up in the state of the mental model. To summarise, in the simplest sense workload manipulations increase the rate at which decisions must be made while SA manipulations effect the timeliness of goal achievement.

7. ACKNOWLEDGMENTS

I would like to acknowledge Gail Grainger, a 1989 DCIEM Summer student, for planning and running the ATC experiment reported in this paper.

8. REFERENCES

1. Yeh, Y. and C.D. Wickens, "Dissociation of performance and subjective measures of workload: Measurement of workload by secondary tasks", Human Factors, 30(1), 1988, 111-120.
2. Hendy, K.C., J. Liao, and P. Milgram, "Combining time and intensity effects in assessing operator information processing load", (submitted to Human Factors), 1994.
3. Powers, W.T., "Feedback: beyond behaviorism", Science, 179(4071), 1973, 351-356.
4. Hart, S.G. and L.E. Staveland, "Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research", *in* P.S. Hancock and N. Meshkati (Eds.), "Human Mental Workload", Amsterdam, The Netherlands: Elsevier Science Publishers, 1988, 139-183.
5. Taylor, R.M., "Trust and awareness in human-electronic crew teamwork", *in* T. Emmerson, *et al.* (Eds.), "The Human-electronic Crew: Can They Work Together?", Wright Patterson Air Force Base, Dayton, OH, USA: US Air Force Wright Aeronautical Laboratories, 1988.
6. Miller, G.A., "What is information measurement?", *in* W.C. Howell and I.L. Goldstein (Eds.), "Engineering Psychology - Current Perspectives in Research", New York, NY, USA: Appleton-Century-Crofts, 1971, 132-142.
7. Hendy, K.C., "An information processing model for workload and performance prediction", *in* K.C. Hendy (Ed.), "Workshop on Task Network Simulation for Human-Machine System Design", Washington, DC, USA: The Technical Cooperation Program, Subgroup U, Technical Panel 7, 1994, 25-45.
8. Hendy, K.C., "Implementation of a human information processing model for task network simulation", DCIEM No 94-40, North York, Ontario, Canada: Defence and Civil Institute of Environmental Medicine, 1994.
9. Kinchla, R.A., "The measurement of attention", *in* R.S. Nickerson (Ed.), "Attention and Performance VIII", Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Publishers, 1980, 213-238.
10. Rasmussen, J., "Skills, rules and knowledge; signals, signs and symbols and other distinctions in human performance models", IEEE Transactions on Systems, Man and Cybernetics, SMC-13(3), 1983, 257-266.
11. Klein, G.A., "A recognition-primed decision (RPD) model for rapid decision making", *in* G.A. Klein, *et al.* (Eds.), "Decision Making in Action: Models and Methods", New Jersey, New York, USA: Ablex Publishing Corporation, 1993, 138-147.
12. Endsley, M.R., "Design and evaluation for situation awareness enhancement", *in* "Proceedings of the Human Factors Society's 32nd Annual Meeting", Santa Monica, CA, USA: Human Factors Society, 1988, 97-101.
13. Powers, W.T., "The fallacy of behaviorism", *in* G. Williams (Ed.), "Living Control Systems II — Selected Papers of William T. Powers", Gravel Switch, Kentucky, USA: The Control Systems Group, Inc., 1992, 233-240.
14. UTP-7, "Minutes of the annual meeting, 21-24 June 1994, Melbourne, Australia", Minutes of Technical Panel UTP-7, The Pentagon, Washington DC, USA: The Technical Cooperation Program, 1994.
15. Selcon, S.J. and R.M. Taylor, "Evaluation of the Situational Awareness Rating technique (SART) as a tool for aircrew systems design", *in* "Proceedings of the AGARD AMP Symposium on 'Situational Awareness in Aerospace Operations'", Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development, 1989.
16. Wilkinson, L., "SYSTAT: The System for Statistics", Evanston, IL, USA: SYSTAT Inc., 1984.
17. Selcon, S.J., R.M. Taylor, and E. Koritsas, "Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation", *in* "Proceeedings of the Human Factors Society's 35th Annual Meeting", Santa Monica, CA, USA: Human Fators Society, 1991, 62-66.

Will the mission workload profile allow effective situational awareness?

S E Jensen,
Human Factors Group
DRA Farnborough, Hampshire,
GU14 6TD
United Kingdom

1. SUMMARY

This paper describes a video-based technique, C-SAW (pronounced 'see-saw'), for developing a time- and task-ordered mission profile of workload with a resolution of as little as 3 seconds real time and much less if used freeze-frame. The rating scale can be based on any uni-dimensional rating scale and can also be used with some multi-dimensional scales. When C-SAW is based on the Bedford scale, the result gives a good indication of the 'spare' attentional capacity which can be devoted to situational awareness. A proposed extension of the C-SAW approach to provide a specifically SA mission profile, based on SWAT, is described.

2. INTRODUCTION

The importance of an extensive and constantly updated mental model of the tactical and overall situation (situational awareness) to operational effectiveness and safety cannot be overstated. Acquiring and maintaining this model requires considerable cognitive and perceptual resources which are competed for by the demands of the priority tasks of flying and weapons management. The availability of increased amounts of information, perhaps from airborne command and control, or from other members of the formation over datalink, is no guarantee that that information can actually be incorporated into the situational awareness (SA) mental model. Indeed, the added task demand of integrating the constant influx of information may simply 'max-out' the crew, leaving even less attentional capacity available to maintain SA.

SA is a difficult concept to pin down; like health, intelligence and workload, it is a complex entity: you know when you have it, but it is not susceptible to direct measurement. One thing is certain, however, the maintenance of an up-to-date SA mental model generates a continual demand for perceptual, central and psychomotor attentional resources, additional to the central tasks of flying and weapon management.

A major thrust of the mission workload research being carried out by the DRA Human Factors Group (sponsored by MOD(Air) Operational Requirements 5), is aimed at the 'spare' attentional capacity available to aircrew for maintaining and updating situational awareness (SA), as part of an overall strategy of attempting to link crew performance to overall operational effectiveness. Part of our initial strategy has been the establishment of the relationship between mission workload and the spare resources available for SA, and to relate these two parameters to the pattern of task demand obtaining through a typical mission.

To apply the results of workload assessment and to draw conclusions on the impact on situational awareness, it is necessary to pin-point accurately the tasks that give rise to workload fluctuations. Established subjective methods, such as the Bedford Scale (1), the Subjective Workload Assessment Technique (SWAT) (2) and NASA's Task Load Index (TLX) (3), relate well to performance, but only for relatively gross coherent 'chunks' of a flight. Computer modelling techniques based on assigning 'workload' values to tasks in the task timeline, while invaluable for guiding an evolving design for a future systems, are inevitably based on some very broad assumptions of the final design and of operator response, and cannot accommodate the range of operator capability and training. Where a simulation or real flight is to be assessed, direct measurement of workload can be made from simulator subjects and it is pointless to risk the compromises of accuracy inevitable in modelling when real subjects are available to operate the real or simulated system. However current conventional subjective techniques are too coarse to guide design detail, so C-SAW has been developed for these later stages of procurement.

3 WORKLOAD AND SITUATIONAL AWARENESS

Merely characterising 'spare' capacity per se is of little use unless it is related to the pattern of tasks being carried out and the progress of the mission itself. (C-SAW) (4) is being developed to provide a time-ordered subjective workload profile with a resolution down to five seconds or less. The profile shares a timeline with a task timeline analysis and mission 'storyline' so that even transient fluctuations in workload can readily be related to their cause. Subjects are asked to recall and rate workload immediately after a flight or simulator run with the aid of video of their cockpit activity. The Bedford

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

scale descriptors, with their emphasis on spare capacity, are used as a basis for the ratings, which are carried out in response to a regular prompt from the computer software. The profile can be generated from a real-time replay, but the option for slow time or freeze-frame replay is available for greater task detail, or as an aid to knowledge elicitation for cognitive task analysis.

C-SAW has shown itself highly sensitive to different attack profiles, between which NASA's TLX and the conventional Bedford Scale were unable to discriminate. It reliably discriminated between two identical attack runs flown against the same target, but with a designation system operating in a slightly degraded mode on one of the runs. The time-ordering method

Background

The specific problem which motivated the development of this technique was the need to assess the demands of single-seat operation of a targeting system originally designed for use by the navigator in a two-seat fast jet. The task requires the pilot to fly a complex attack profile using the head-up display (HUD) symbology, overlaid on the head-down display (HDD) of the targeting device. The targeting display has to be monitored throughout flight to ensure that a marker remains on the target. Although the DRA test pilots had established that single-seat operation of the device was possible, they were aware that they were working at the limits of their attentional resources and were very conscious of the safety pilot with them, who, while not assisting them with the task, would not allow them to endanger the aircraft and themselves. In order to make a more quantifiable assessment of workload, a method of capturing the momentary workload levels through the attack run was necessary.

What was needed was a method of visualising a detailed time-ordered profile of flight workload that could be read across to an equally precise task timeline. An attack run is very eventful and the established subjective and physiological techniques for workload assessment were quite incapable of giving the resolution needed.

Developing Continuous Subjective Assessment of Workload - C-SAW

Taking workload ratings from aircrew during an attack run would be impossible, not to mention dangerous, so the technique was based on a recapitulation of an in-cockpit film of the attack run. Preliminary studies comparing in-flight commentary of subjective workload with the post-flight C-SAW suggested that the aircrew could recapitulate their subjective experience of workload quite consistently, provided the film was viewed immediately after they had landed. While the film played at normal speed, the subject pressed one of ten keys on a keypad, corresponding to the descriptors of the Bedford Scale (1) in response to a prompt from a computer. The Bedford Scale was used here, as it was necessary to use a uni-dimensional scale and the Bedford addresses workload and 'spare' resources specifically, so that some conclusions on the potential for situational awareness can be drawn. In this case, the Modified Cooper-Harper Scale (5) was not appropriate, as it is directed more at errors and the interface design. C-SAW can be used with any uni-dimensional rating scale (not necessarily just workload) with a range from 0 to 100 so that adapting it for use with the Situational Awareness Rating Scale (SART(6) or a specifically developed SA scale should be relatively straightforward.

Surprisingly, initial 'pilot' runs of C-SAW showed that the subjects could respond reliably to prompts as frequent as every 3 seconds, provided they were not required to maintain this rate for too long. An attack run is normally complete well within two minutes, and the aircrew have little difficulty in maintaining the 3-second input rate for this time. The software collects the data in the form of ASCII text files which can be read into any suitable software package and printed as a bar-chart or graph against the timeline.

Before the experimental flight, a theoretical task timeline is established by consultation with the aircrew, and for each individual flight an accurate timeline is calculated from the video film. The workload ratings and individual task timelines are then combined to provide a flight workload profile and task description with a common timeline.

The data in Figures 1 and 2 are illustrations of C-SAW output from two attack runs on the same target using the same attack profile, but with the system operating fully in Figure 1 and in a degraded but still operationally effective mode in Figure 2. The C-SAW output can be seen clearly to respond to the different workload levels of the two conditions, and the associated on-going tasks can be read off from the task timeline below. NASA Task Load Index (TLX)[3] overall rating and the conventional Bedford Scale ratings for the attack runs are also shown on the C-SAW chart.

C-SAW has been used in flight trials and in simulations, and has achieved a very high face validity. Formal validation is planned in the relatively controllable simulation environment, both with full fidelity simulation and in a multi-workstation computer-based tactical simulator. The criteria being used are test/retest consistency; both for individual subjects and for differences between subjects. Comparison with established techniques is difficult, as

Figure 1 C-SAW Profile of attack run with fully functional designator

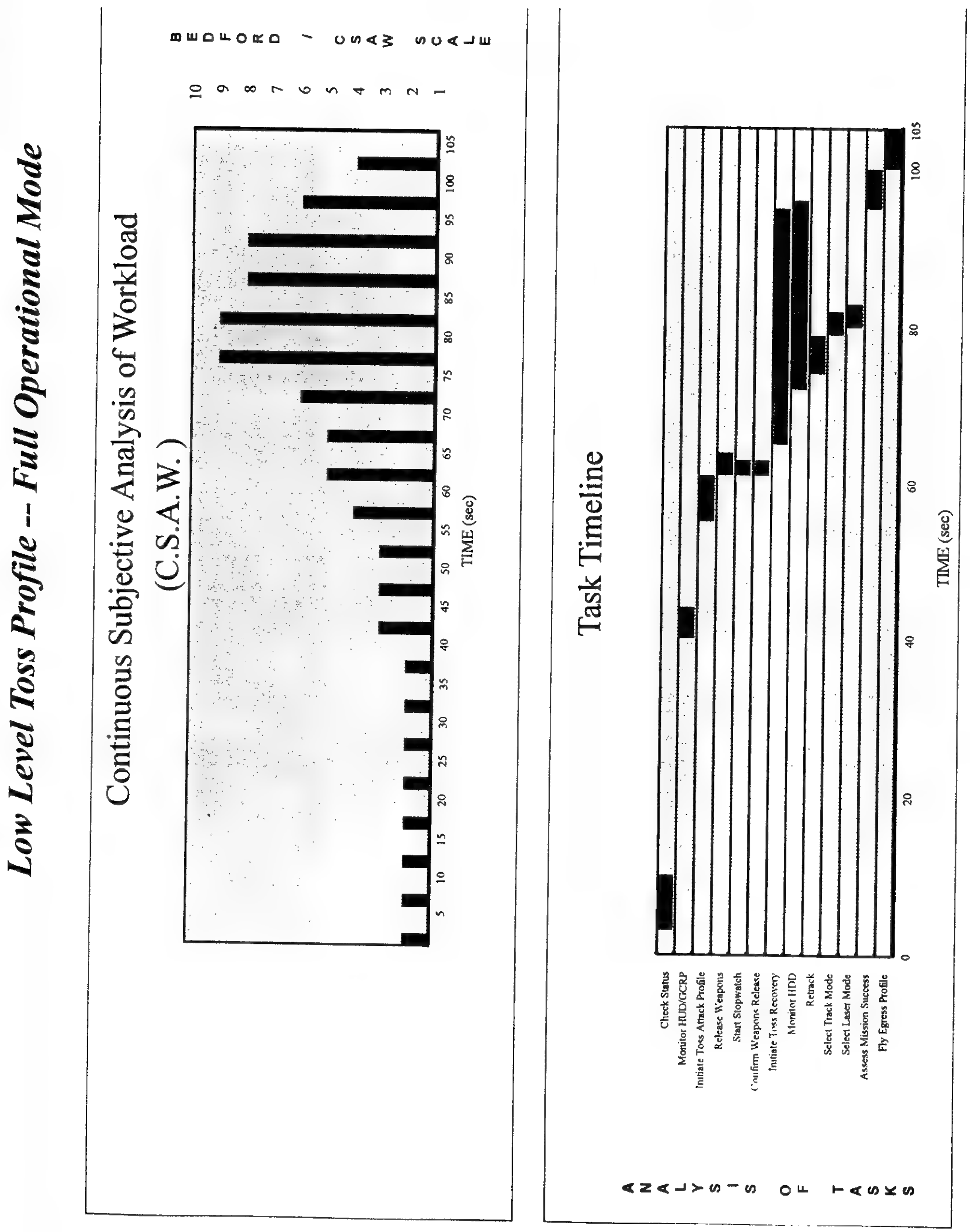
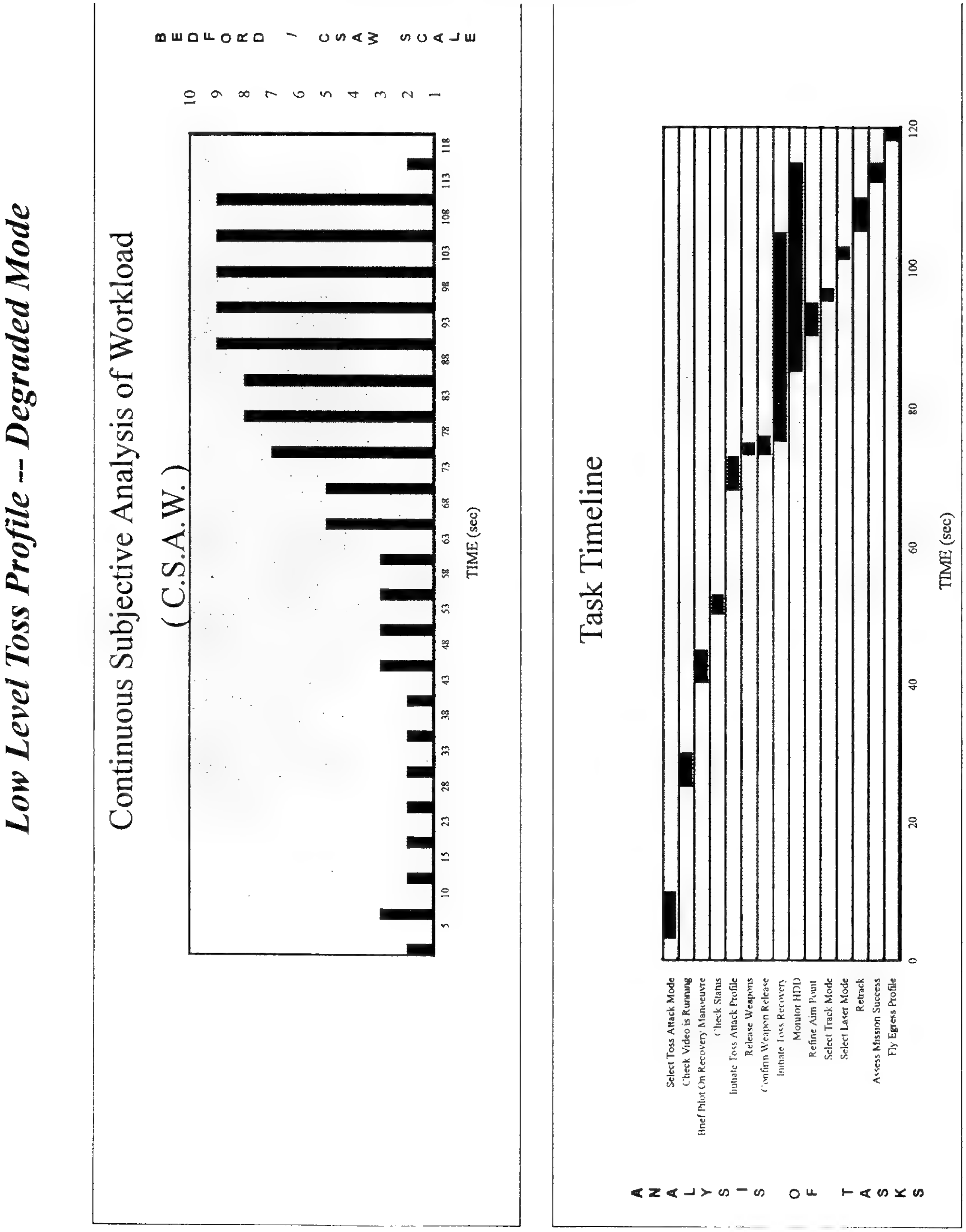


Figure 2 **C-SAW Profile of attack run with designator degraded**



these give only an overall rating for the whole time-period rather than a time-ordered profile, as is illustrated in the two Figures shown below.

The C-SAW technique can also be used in a freeze-frame mode for a very detailed investigation, perhaps when a particular display or manoeuvre is being studied, or where an area of interest identified by the initial 'real-time' assessment needs to be studied in greater detail. A separate version of the software has been developed for this purpose, and it is this version which we are developing for SA studies.

The approach we are assessing at the moment is to extend the basic C-SAW approach by basing it on the dimensions of the 3-D SART (attentional demand, attentional supply and understanding). We will have an objective basis for the attentional demand rating in the task time-line analysis; C-SAW itself, with the Bedford scale descriptors, will supply a time-lined measure of attentional supply. 'Understanding' is the remaining measure and we plan to run the video recording and C-SAW software a second time, with the subjects giving a subjective rating, possibly on the low/medium/high basis used by Taylor (6).

Validation of the C-SAW approach for SA (Continuous Subjective Assessment of Situation Awareness, C-SASA) will be more tricky than the same process for workload, as it requires aircrew to recapitulate a state of objective knowledge, which itself will be affected by hindsight. Studies about to begin at DRA, and aimed at evaluating the effectiveness of visually coupled systems in attack helicopters using a CGI simulation, will enable us to validate C-SASA against objective SA performance criteria, by injecting artificial situation changes and asking for very frequent situation reports over the simulator's 'radio' link. Subject experts acting as observers and the subjects themselves will also be able to examine the C-SASA output and assess its accuracy.

To return to C-SAW, further work will extend the C-SAW method to the investigation of particular types of workload, such as auditory/verbal workload when communications are being investigated, or aspects of cognitive workload such as short-term memory during complex decision-making. Other areas where it is hoped to assess C-SAW's potential include over- and underload in civil aviation and workload in non-aviation environments, such as process control.

References

1 Roscoe, A. H., (1987) 'Introduction to practical assessment of pilot workload' in Muir, H.C., Roscoe, A.H., (eds) *Proceedings from the Symposium on the Practical Assessment of Pilot Workload* AGARD, Paris.

2 Reid, G.B., Shingledecker, C.A., Eggemeier, T.F. (1981) *Application of conjoint measurement to workload scale development.* Proceedings of the Human Factors Society 25th Annual Meeting.

3 Hart, S, Staveland, L E, (1988) 'Development of the NASA TLX: Results of empirical and theoretical research' in Hancock, P A, Meshkati, N, (eds) *Human Mental Workload* North Holland, Amsterdam.

4 Jensen, S. Developing a flight workload profile using Continuous Subjective Assessment of Workload (C-SAW), 1994. Proceedings of the 21st. Annual Conference of the Western European Association of Aviation Psychologists, Trinity College Dublin. (in press)

5 Wierwille, W. W., Casali, J. G. (1983) *A Validated Rating Scale for Global Workload Measurement* Proceedings of the Human Factors Society 27th Annual Meeting.

6 Taylor, R M, 1989, *Situational Awareness Rating Technique, (SART); the development of a tool for aircrew systems design.* in: Proceedings of the AGARD AMP Symposium on situational awareness in aerospace operations. Copenhagen, Denmark, 2-6 October 1989.

THE FLIGHT SAFETY OFFICER'S PERSPECTIVE

WING COMMANDER N B SPILLER RAF
Headquarters Strike Command
RAF High Wycombe
Buckinghamshire
HP14 4UE
England

SUMMARY

1. RAF aircraft accidents in the last 2 years have included a Hercules in Scotland, where all 9 on board were killed, 2 Harriers, one on operations over Iraq and one on a training flight in England, and a Tornado F-3 and its crew off Cyprus. The RAF also suffered 2 mid-air collision accidents, one involving a Tornado GR-1 and a civilian helicopter and the other 2 Tornados in formation over Canada. Human factors and the maintenance of situational awareness were a feature in all these accidents. This paper attempts to identify some of the human factors linking these accidents.

INTRODUCTION

2. It is a great pleasure to be invited once again to address an AGARD audience. The last occasion was some 7 years ago when I was a younger staff officer in a bustling NATO HQ, the Berlin Wall was still very much in place and I gave the Operator's Perspective. This time, I come to you as an older and, hopefully, wiser Command Flight Safety Officer to give a flight safety perspective.

3. I note that my position in the order of presentations has also changed over the years. Last time, I was one of the first speakers and was then able to sit back and listen to those far more qualified than I analyze the problems that I had identified. This time, I have no such luxury. The experts have all spoken first and I am left wondering what there is new for me to say?

4. Let me try. My aim today is to review some fairly recent RAF accidents - sadly, we have not been short of examples - and attempt to identify the threads in these accidents that led to failings in situational awareness. Inevitably, I will be concentrating on peacetime training; however, I would also like to highlight some operational challenges that lie ahead. My only qualifications for speaking to you today are some 25 years tactical flying and more time tramping the ground of lonely crash sites that I would have wished.

HERCULES ACCIDENT

5. Fortunately, it is not often that most of us lose a big aircraft but in 1993, to my certain knowledge, the Canadians, the USAF and ourselves all lost a Hercules C-130. I would like to take you through our accident in some detail as it has lessons for us all.

6. The aircraft was conducting tactical reinforcement training. It was being flown by an operational crew but with a screen pilot and navigator also on the flight deck to supervise and monitor the flight. The low level portion of the sortie was to include a practice parachute drop. The DZ chosen for the practice drop is shown on this map.

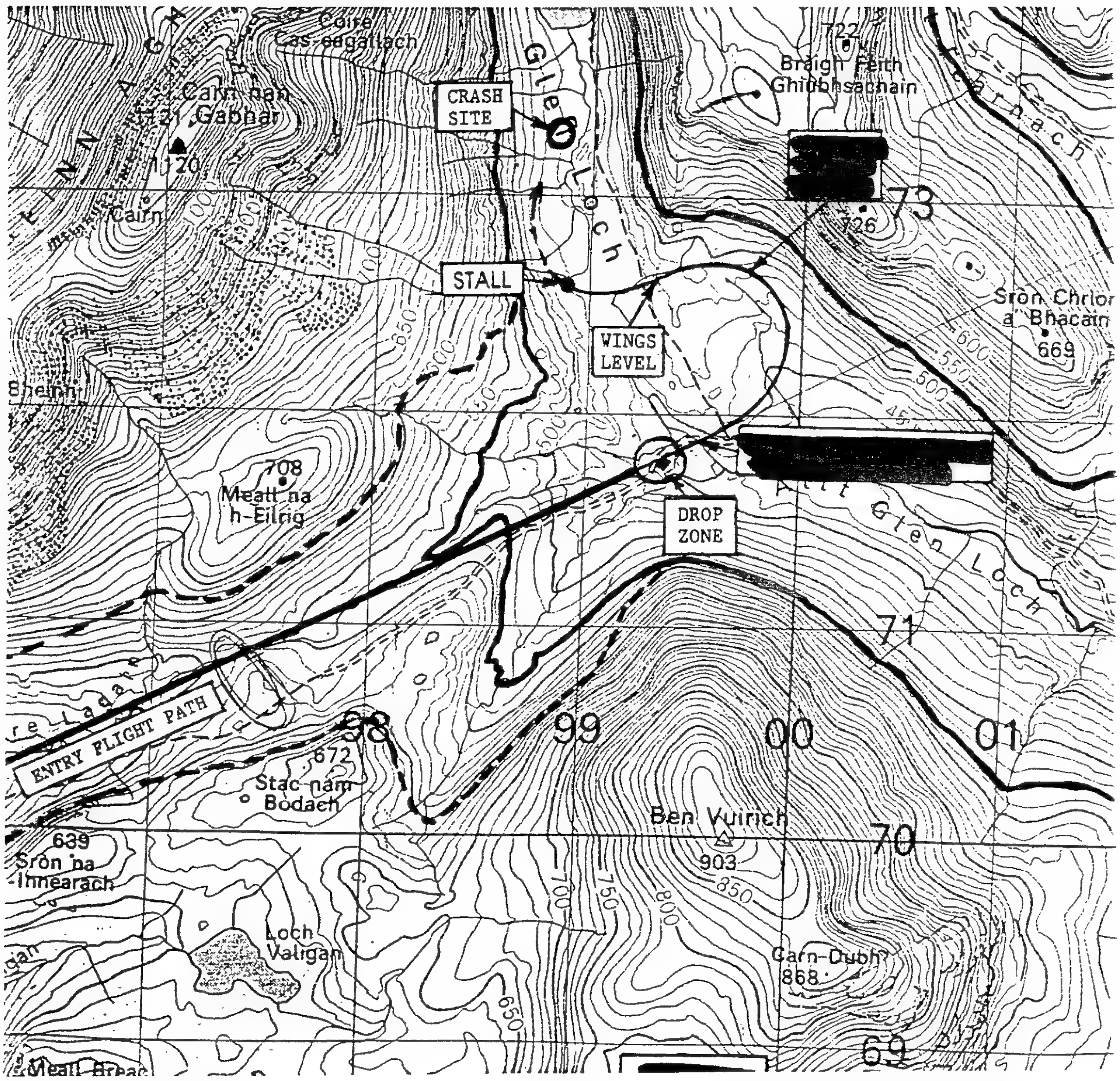
7. The crew planned to approach the DZ from the South West down this valley and to exit the DZ to the North up this valley. This is what the DZ looks like and, as you approach it from the South West, if you glance left to look up your exit route this is what you see. It all looks fairly benign. The problem comes when you are over the DZ as the view to the North then looks like this. A look at the map shows their final route and the cause of their problems. This rock buttress meant that, if they were to maintain 250 ft, they had to turn nearly all the way onto West before being able to reverse onto North. You will see that they are now facing an almost vertical rock face ahead of them.

8. All is not yet lost - but nearly so. If the crew had realised their predicament, they could have continued their left turn and escaped up the valley that they had just come down. That would have shown real situational awareness. In the event, they reversed their turn to follow the plan. The Hercules takes about 9 seconds to roll from 45 degrees of bank one way to 45 degrees of bank the other way - and all the time this mountain is getting closer and more threatening. The crew were left with the choice of flying into this mountainside or desperately trying to haul the aircraft round the corner.

9. The aircraft struck the ground 28 seconds after passing the DZ with 10 degrees of right bank and the nose 15 degrees above the horizon. Forward speed was 87 kt and the rate of descent was 1,600 ft per minute. All 9 people on board were killed.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

HERCULES ACCIDENT - FINAL FLIGHTPATH.



10. There are several points that I would like to highlight about this accident. Firstly, this was an operational crew conducting reinforcement training. They were not a student crew that needed close supervision. Moreover, they were being overseen by an experienced screen captain and navigator. The sortie that they were flying was a routine training sortie using a DZ that had been used many times before. Do you sense the warm comfortable feeling of complacency? As our Air Force draws down and becomes less mobile, people are tending to stay longer in one place, flying the same routes and profiles. Several of our recent accidents have occurred to people long established in apparently safe posts who have become complacent and, so, not taken the necessary degree of care.

11. Let's turn our attention to the crew. I expect that the operating crew, who were being monitored by the screen crew, would have wanted to do well, be seen to be tactical. If they had simulated their drop short of the DZ or from a greater height, many of the problems of their plan disappear. But the crew had been given this DZ by the instructors. There is a logic path in their brains that says, "They would not have given me this DZ if I cannot simulate a drop on speed, exactly over the DZ and at 250 ft". And remember, they want to impress. They want to be exactly on target, on time, on speed and on height. We should not be surprised by this - we recruited them for just these characteristics. Air power, by its very nature, has to be aggressive. We want young people who will go out and actively engage the enemy. This, by the nature of the medium we fly in, means risk. It is these same young people who will want to drop exactly over the DZ, on speed and on height. Moreover, in their search for airborne perfection, they may be prepared to take risks that other lesser mortals might not deem prudent.

12. You will all, I am sure, be more familiar than I am with Crew Resource Management, LOFT, and the other cockpit management initiatives. The plan for this sortie was undoubtedly flawed - but nobody spoke up. I sometimes wonder if the co-pilot on this sortie, had he seen the problem, would have spoken up against his own captain and a screen captain and navigator. Like it or not, our aircrew operate within a disciplined military structure and the consultative approach encouraged in civilian flying is not so easy for us to cultivate. This is a particular frontline military problem. How do we persuade our aircrew to question any potentially unsafe act in peacetime and yet be prepared to face without question the risks associated with live operations. Moreover, live operations that risk becoming increasingly deadly and, in future, may not involve the defence of your own country with all the emotion and motivation that this can generate. But which involves so called peacekeeping missions in an area of the World that you may not even have heard of a few months before.

HARRIER ACCIDENTS

13. Let me turn to another of our accidents. This time a Harrier that flew into the ground while evading at low level. The RAF chooses to do much of its tactical training at low level for sound operational reasons - but it is a regime requiring high levels of situational awareness. On this occasion, the pilot concentrated on the attacking aircraft and his wingman to the exclusion of all else and flew into the ground. He was not inexperienced. He was a USMC pilot on exchange with over 1000 hours on AV-8B - but most of his experience had been at medium level - and for a few critical seconds on this sortie he forgot that he was at low level. There are undoubtedly additional pressures for a pilot on exchange with another air force. A desire to do well, to bring respect and honour to your own Service and Country. Statistically, in the RAF, an exchange pilot flying our aircraft runs twice the risk of having an accident as does his RAF counterpart. Twenty four accidents and 12 fatalities in the last 25 years - that is a high price to pay for representing your Country in Peacetime.

14. Another Harrier accident that we had recently occurred on operations over Northern Iraq. The aircraft was refuelling from a VC-10 tanker when the engine changed from digital to manual fuel control. This was possibly caused by a cognitive failure by the pilot but the result for him was to go suddenly from the benign world of VMC flying with all systems operating to IMC with no engine and descending towards mountains that rose to over 8500 ft in the clouds below - and that he knew to be Iraqi mountains. Along with his engine went many of his electrics - including his computer control for the engine and all his easily interpreted engine indications in the Head Up Display. He was left with some mechanical digital indications that, in the stress of the moment and with parameters changing fast, must have required time and concentration to interpret.

15. Aircraft have become very much more reliable over the years and our aircrew today are not as used to dealing with in-flight emergencies as were their forebears. The simulator can go someway to compensate for this but, no matter how hard you try, simulators cannot engender the same levels of fear. Fear and confusing or complex information can lead to paralysis of the mind or a desire to rush into the drills - almost any drills - to sort the situation out. This, in turn, leads to inaction or the wrong actions being carried out - either of which risks losing an aircraft. The influence of fear on human performance is equally important in combat. How do you know who is going to perform well when they are frightened? Our experience in the Gulf Conflict suggests that it is not always the people that we think it might be. A look at the biographies of some of the great fighter aviators from the past shows that several of them almost

certainly would have failed our modern selection procedures. I just ask - can we measure beforehand how a person will perform when frightened? If we can, this should certainly be a part of our selection procedure.

TORNADO ACCIDENTS

16. Let me turn to another accident. We also lost a Tornado F3 and its crew last year. The crew were conducting air-to-air firing to the South of Cyprus. They finished their last firing pass, turned towards base at 5000 ft and put the aircraft into a 2 to 4 degrees nose down descent. A short time later they swept the wings fully aft, and selected full afterburner. They continued to accelerate and descend until they hit the sea at some 590 kt and 6000 ft/min. I don't know how many of you are familiar with Cyprus but the conditions at the time were not unusual for much of the Mediterranean - hazy, indistinct horizon, fog that looked like cloud to within a km of the crash site and a smooth, glassy sea. We had all the visual illusion ingredients for the pilot to mistake his altitude - but why did not he, or his navigator, check the height. In fact, there was virtually no talk between the crew. There were no post-gunnery checks, no height checks and no recovery checks. Why?

17. Analysis suggests that they had flown together so often that they had, perhaps, developed too much trust for each other's abilities. We do not put 2 people in a fighter aircraft only to monitor each other's actions. In combat each has a vital contribution to the success of the mission. But on recovery, when the workload is not high, we can, and do, use the extra person to monitor and provide the vital safety cross-checks - yet, on this occasion, they trusted each other too much to bother. A fatal mistake.

18. There were other factors in this accident. As we have drawndown our forces since the fall of the Berlin Wall, as have most NATO countries, we have not reduced at the same pace our commitments. Indeed, there are new commitments for our squadrons like Partnership for Peace - the NATO liaison with former Warsaw Pact countries and, of course, our operations in the Middle East and the Balkans. The result of all this is that our Squadrons are actually working harder than they were during the Cold War. This squadron had had a demanding year, including many overseas detachments and preparations for a demanding 4 month operational deployment. Cumulative fatigue is difficult to identify and even more difficult to measure - we will never know if it was a factor in this accident.

19. Another issue was the pilot's state of mind. He had had an accident some 7 months before in which he had ejected successfully from a Tornado F3. The cause of the emergency was a technical failure but his mis-handling of the emergency had caused the aircraft to crash. In fact, he had shown the classic haste that I mentioned earlier in rushing into inappropriate actions that, ultimately left him with no engines, no speed, no height and no ideas - apart from to rely on Martin Baker. He was still awaiting disciplinary action for this earlier accident. Was this playing on his mind? Moreover, if he had been found wanting once, should we have continued to employ him as a pilot?. Can you re-train someone to react correctly the second time, or are they a lost cause? You will be able to answer this better than I can.

20. Finally, I know that you are all asking - what about the radar altimeter? That should have given them a clear audio and visual warning of their approach to the sea. Yes, it should. But the designers of the aircraft had decided that the nose wheel steering failure and radar altimeter low altitude warning could use the same 600 KHz horn, so the standard procedure on recovery was to switch the height warning bug to zero so that the nose wheel steering warning would be available on landing. The crew were denied a vital height warning that could have saved their life by poor design. We cannot afford to make mistakes like that in the future.

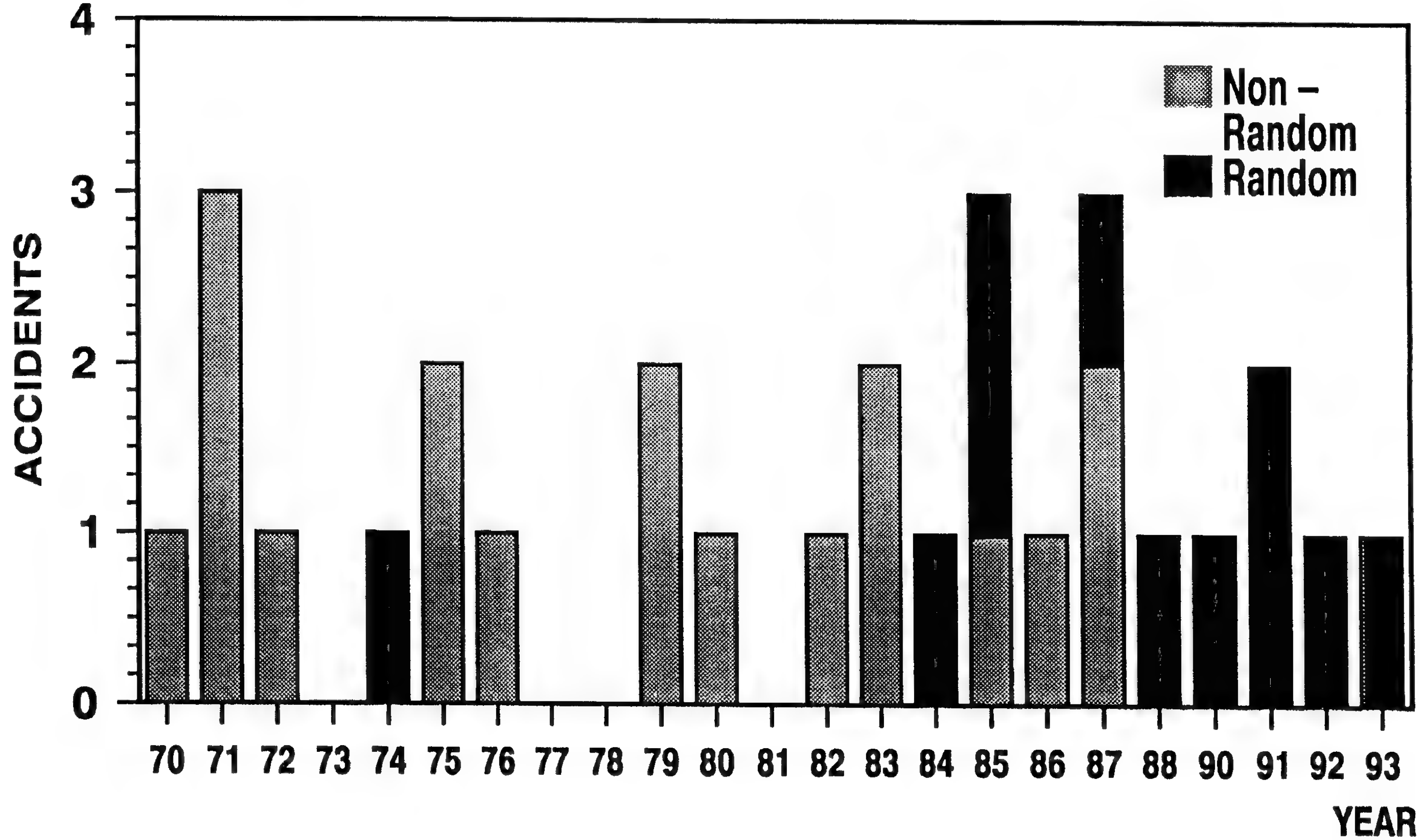
MID-AIR COLLISIONS

21. Let me talk briefly about mid-air collisions. Last year, we had an accident over Canada where the No 4 in a 4-ship collided with his No 3. It would be nice to say that they were working at the leading edge of tactical training on an exceptionally challenging mission. Unfortunately, that was not quite the case - they were transitting back from Alaska, where they had been on Exercise Cope Thunder, behind a Tristar Tanker. The tanker called and turned left 20 degrees. The 4 pilots all put the new heading into their autopilots and, shortly afterwards 2 aircraft collided. One crew ejected and one badly damaged aircraft was lucky to make it to its diversion airfield.

22. A few days before, these crews had been conducting coordinated day and night attacks at heights down to 100 ft with a fighter bounce overhead and a realistic SAM threat on the ground. They were not incapable of flying accurate formation. Of course, as a breed, aircrew are drawn towards the more exciting training sorties. Their very nature makes them shy away from the repetitive, the routine - but formation on autopilot and no sense of self-preservation when floating along in company with 4 other aircraft - Amazing!

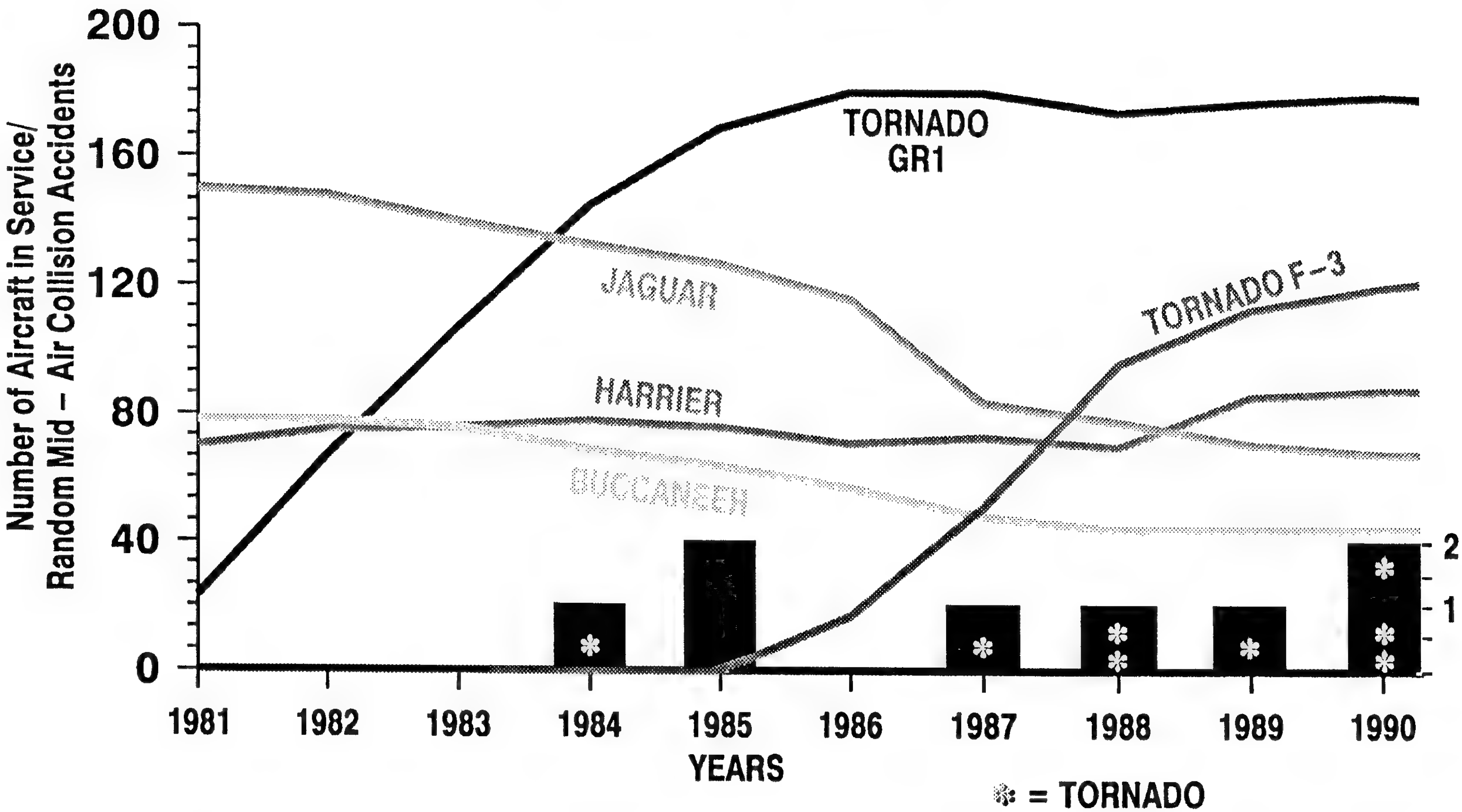
23. It is my view that a factor in this accident may have been the search for financial savings in our training system combined with our more hectic squadron routines. In the stable days of the Cold War, training profiles were relatively unchanging and, once on a squadron, there was a fairly fixed work up routine to

MID–AIR COLLISION ACCIDENTS 1970–1993



19502

AIRCRAFT STRENGTHS/RANDOM MID – AIR COLLISIONS AGAINST TIME



19501

Combat Ready status. NATO even checked on your progress at the annual TACEVALS. The result was that, at almost any stage, you knew a pilot's ability level - what he had done and could do safely and what he had not done. Recent events have changed all that. The need to prepare for live operations means that some of our crews are far more experienced than we would expect at certain disciplines but, equally, have not done to the same standard things that we might have taken for granted 5 years ago. Here lies the danger - and accidents while conducting apparently routine tasks, like this transit across Canada, could be the result if we are not careful.

24. We also had a mid-air collision accident in 1993. This time between a Tornado and a civilian helicopter. The Jet Ranger Helicopter crashed immediately after the collision, killing the 2 crew on board. The Tornado was badly damaged but managed to land safely.

25. There was the, perhaps, predictable public outrage following the accident. The public saw this as a civilian helicopter going about his lawful business and being knocked from the sky by the RAF. The reality, of course, was not quite like that. The detailed investigations that followed the accident determined that neither crew had had any real chance of seeing the other aircraft in time to avoid a collision. This did not stop the Press insisting that something should be done and questioning if we could continue to operate safely at low level.

26. Mid-air collisions remain a risk for every air force. We, at Strike Command, in the light of this accident conducted some research on our mid-air collision statistics. This slide shows the number of mid-air collision accidents that occurred to our frontline aircraft between 1970 - 1993. It does not include our training aircraft or the Red Arrows. The green bars indicate collisions that occurred between aircraft in the same formation or where the pilots knew of the existence of the other aircraft. The accident over Canada, for instance. The red bars show the random conflictions between aircraft that were in different formations and where neither crew was aware of the other aircraft.

27. These mid-air collision accidents cost us 41 RAF aircraft, 4 NATO, Army or Navy aircraft, and 3 civilian aircraft. They also cost 51 people their lives. The thing to note about this graph is the change in the cause of most of our mid-air accidents in the mid eighties. There are various reasons for this change - more aircraft, both civilian and military, using the airspace - fast jet aircraft replacing our older and slower bombers. This graph shows our frontline strengths over the period. You will notice the rise in the number of Tornado that corresponds closely with the increased level of risk. We are also operating our aircraft differently, both tactically and in our systems management - more time heads in looking at EW equipment, inertial nav to look after the navigation so that you spend less time heads out looking for turning points - I could go on.

28. We are taking positive steps in the RAF to address the mid-air collision risk. Work by John Chappelow, who I know will be known to many of you, has shown that black aircraft provide better contrast and are easier to see. So, we are painting our Hawk training aircraft black. Operational aircraft will not be painted black but will remain in their present camouflage so that we can meet our operational commitments. However, for our operational aircraft, and in particular the Tornado, Harrier and Jaguar, we are developing a Collision Warning System. The technology demonstrator has already flown successfully and we are now looking at the problems of integrating the equipment into fighter aircraft. The idea is that the Warning System will give information on aircraft approaching your aircraft in the form of both audio and visual alerts. The system will not provide advisories as does TCAS but will work at speeds appropriate to fighter operations, which TCAS does not, and will also not warn on other members of your formation, which TCAS would.

THE HUMAN FACTOR

29. Let me try and bring together the various threads that I have alluded to in my discussion of these accidents. My first point, I think, is that, as a pilot on a squadron, and particularly as a supervisor or a squadron commander, you can fairly quickly tell who has innate ability at maintaining situational awareness. What's more, sadly, it is not always the majority of the squadron - yet it needs to be if we are to maximise our combat capability. Research has shown that, traditionally, about 5 percent of fighter pilots obtain about 90 percent of the kills. Our aim has to be to increase that 5 percent so that as a force we are more effective.

30. Now, if fellow pilots on frontline squadrons can identify their colleagues with good situational awareness, is there not some way that the selection and training process, with its wealth of highly qualified psychologists, could not also identify these individuals, but earlier so that we only train the really competent individuals who will uprate our combat power?

31. You will all remember the talk of the 'Right Stuff'. But what is the 'Right Stuff'? Last time I was at one of these gatherings, it was suggested that the solution lay in a 2-seat aircraft with a stable extrovert in the front seat and a stable introvert in the back seat. Maybe, but we have to acknowledge that for all

European Air Forces, the future lies predominantly in single seat aircraft. Eurofighter 2000, Rafale, Grippen, F-16, F-18 and their follow on - these are the aircraft of the next century.

32. Moreover, the air war is becoming ever more complex. It is relatively easy, if not cheap, to build a multi-role aircraft. It is far more difficult to maintain current a multi-role pilot. I would argue that among the most difficult sorties demanded of our young aircrew are night ground attack below safety altitude and, on the air-to-air side, the sorting, targeting and maintaining situational awareness during the 4 versus many scenarios. One NATO squadron commander suggested to me that the maximum currency window for staying at full proficiency during complex air-to-air missions was about 2 weeks. Flying hours are at a premium and, for some Air Forces, it is proving difficult enough in peacetime to find sufficient high-grade training to maintain proficiency in one discipline - never mind 2 or 3.

33. The solution for many Air Forces is to adopt a training cycle, bringing crews to peak proficiency at the various disciplines in turn throughout the year, and relying on any necessary top-up training being available before having to conduct operations. This is sensible, but it still leaves the question of what can we do to help this potentially overworked pilot. First, good, readily interpreted information is vital. The addition of JTIDS to our aircraft has had a profound effect on the ability of our crews to maintain situational awareness. Indeed, I would argue that JTIDS will have as significant an effect on the successful prosecution of the air war as did the introduction of airborne radar. But the system has to be user friendly. It must be straightforward to operate. Crews cannot afford to spend time or concentration operating the system. We need to channel all their computing power into fighting the war not operating the equipment.

34. Next, and returning to my flight safety theme, if our pilots are to be so immersed in the complexities of fighting the air war, then the aircraft must be so simple to operate that it can be done without thinking. No more interesting handling qualities like the Phantom, no more complex fuel contents indications that require time to assimilate, and no more different stages of software, with different buttons doing different things across the fleet. What's more, as we are asking so much more of our aircrew, we need to think more about what I call 'Automatics for Life'.

35. A glance at our losses during the Gulf Conflict will show what I mean. Not all our losses over Iraq were to enemy fire. Some aircraft just flew into the ground while avoiding the ground defences or through distraction. If we are to ask our pilots to prosecute night low level using NVG and FLIR, we need to look also at giving them a system that will stop them flying into the ground. Similarly, if they are to undertake challenging air combat scenarios, we need automatics that will stop them burying the nose so deeply that the aircraft cannot be recovered. I could go on. The British Aerospace TERPROM system, that is being fitted to USAF, US Reserves, Norwegian, Belgian, Danish and Taiwan F-16, provides a good starting point (and no, I'm not on their sales team) but it still requires the pilot to interpret the indications in the HUD and action the voice warnings. Why not go a step further and have the system identify the danger, take control of the aircraft and fly it clear of the ground then, if you like, tell the pilot why it did it. Aviation history is littered with dead crews who chose to ignore a ground proximity warning.

36. Finally, in all this, we need to remember that our basic ingredient, man, has not developed as fast as has technology in the last 50 years. He is basically the same raw material that cowered in the trenches in 1915 and clawed his way into the World's fist dogfights alongside Richthofen, Mannock and the other aces of WWI. Man remains an intriguing mixture. He can demonstrate great acts of sacrifice and bravery. Yet, as a species, he is generally averse to the unexpected and the unknown. He fears failure. He is vulnerable to the shock of battle, and he has finite reserves of energy. In all, he can become mentally and physically exhausted surprisingly quickly. Indeed, one could argue that, were we to set out to design a species for air fighting, man would not even make the fly off. But he is all that we have. Our job remains to make him as effective and safe as possible. Alongside the challenges of technology for the next century, we have to remember that somewhere in the system we have to put a frail and simple human being.

Reaction Time and the EEG Under Hyperventilation

Dr. Wolfgang Freund, Maj. MC
German Air Force Institute of Aviation Medicine
Head: General (MC) Dr. Daumann
82242 Fürstenfeldbruck, Germany

Summary:
EEG recordings at rest and during controlled hyperventilation (HV) with simultaneous recording of reaction time measurements takenfrom 59 pilots and candidates in the German Air Force Institute of Aviation Medicine (GEAF IAM) are used to investigate
1. the correlation of visible EEG changes during HV and changes of cognitive functions and
2. the feasibility of two different methods of measuring the CO_2 (infrared absorption gas analyzer vs. solid body ceramic transdermal electrode) during this experiment.
Under HV the CO_2 decreases from a median of 37 at rest to 22mm Hg during HV. The amplitude of the EEG activity increases as the main frequency decreases. No paroxysmal activity is recorded. Reaction time increases markedly under HV (from 184ms to 226ms).
Surprisingly, no correlation shows between EEG changes and reaction time.
Both methods of CO_2 measurement prove their feasibility, while the gas analyzer is faster and easier to operate.
The implications of these findings are discussed.

Introduction:
The presented study aims to explore the underlying factors that contribute to our evaluation of EEG readings of candidates for military flying duties.
In these EEGs we look not only for signs of epileptic activity but also for signs of abnormal excitability.
This is done so because
a) an aviator´s CNS functions should be stable even under adverse circumstances (maximal excitation is epileptic activity), and
b) some EEGs show changes in their buildup that resemble those seen in coma, leading to the suspicion of impaired CNS function under stress.

By selecting stable candidates, future disruption of situational awareness through impaired CNS function under stress is brought to a minimum.

To establish common ground, the following definitions are given.

Situational awareness (SA)
Situational awareness is the interaction of a person´s knowledge and perception to form an understanding of the current situation (Vidulich et al, 1994).

Situational awareness is understood (Navathe and Singh, 1994) to consist of two subsets: Spatial Orientation and Geographic Orientation. Loss of Situational Awareness may therefore be due to Spatial Disorientation (Physiological limitations) or Geometric Disorientation (lack of skills or training). If neither Spatial Disorientation nor Geometric Disorientation is found, loss of Situational Awareness must be due to psychological factors.

Spatial Disorientation (SD)
SD (physiological limitations, according to Navathe and Singh, 1994) is caused by the combined physiological and pathophysiological effects of external stressors such as G-load, temperature, hypoxia, pain, workload etc. and individual stability in this unphysiological environment.
Wellknown examples are gravitational loss of vision or gravitational loss of consciousness which are caused by the g-force induced decrease in cerebral blood flow (Njemanze et al, 1993) and individual susceptibility.

Selection of candidates at the German Air Force Institute of Aviation Medicine (GEAF IAM):
Candidates who apply at the GEAF IAM are already selected by the institutions that perform the general selection for the German armed forces.
On their first day at the GEAF IAM the candidates are selected and graded at the Department of Aviation Psychology.
The resulting gradings have been validated by a follow-up study that showed probabilities to successfully pass the screening for military flying duties: Candidates graded A, B, C and D had a 82%, 71%, 62% and 40% chance to succeed (Hoffelt and Gress, 1992).
After this psychological testing the candidates are thoroughly examined at the Department of Medicine. Part of the medical department is the neurological and psychiatric specialists group, where the candidate has to undergo a neurological and psychiatric examination as well as an electroencephalogram (EEG) and Visually Evoked Potentials (VEP) (Freund, 1994).

The EEG under hyperventilation:
The EEG electrodes are placed according to the 10-20 System and the recording is done with a 24 channel Schwarzer EEG ED-24.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

Two methods of activation are used: First there are four minutes of hyperventilation and then there are about four and a half minutes of photostimulation (which is disregarded in this paper).
Hyperventilation induces an acute ischaemic hypoxia of the brain: The hypocapnia leads to cerebral vasoconstriction, thereby reducing the cerebral bloodflow, and alcalosis with higher affinity of oxygen to hemoglobin (Kenealy et al, 1986; Adler, 1991). This combined (vascular and cellular) hypoxia leads to both decreased cerebral energy metabolism and lactacidosis (Adler, 1991), differing in result concerning the EEG changes (Van der Worp et al, 1991) from sole hypoxia.
The effects of hypocapnia on EEG are amplified by hypoglycemia (Sieber et al, 1992).
In the spontaneous EEG, HV leads to a reproducible decrease in alpha and beta activity and an increase of slower activity (Kraier et al, 1988) that may mostly be due to the cerebral hypoxia (Adler, 1991). These effects seem to decrease with age (Yamaguchi et al, 1979; Konishi, 1987).
Cerebral vasoconstriction alone (induced by indomethacin) produces only a slowing of the alpha band without effects on theta or delta activity (Kraier et al, 1992).

Figure 1:

HV-Score (revised)

Amplitude	max HV/standard	<=1 1-1,5 1,5-2 >2	0 1 2 3
Abnormality under HV (df>2Hz)		>2min <2min <1min	1 2 3
Quantity of Abnormality/ General Alteration		<10% 10-50% >50%	0 1 2
Onset		gradually sudden	0 2
Normalization		during HV <30 sec <1 min <2 min >2 min	0 1 2 3 4
Focal Alteration	inconstant/exclusively occipit.	constant	2 4
Alterations specific for epilepsy			10
Sum regarded to be pathologic if exceeding 10 points			

All in all, the effects of hyperventilation lead to a metabolic encephalopathia. So the resulting EEG-changes should be similar to those reported for encephalopathia (Velho-Groneberg, 1995): Diffuse slowing and reduction of the alpha band with a consecutively more dominant theta and delta band.
The EEG in the GEAF IAM is examined visually. Regarding the HV-buildup a HV-Score (modified from Glaser and Freund, 1994) as shown in figure 1 is used.

Candidates showing high excitability under HV (HV-Scores> 10) are rated unfit for military flying duties.
This is done for safety reasons (maximum excitability leads to an epileptic seizure under HV) and for reasons that were mainly unproven hypotheses.
Becoming aware of these shortcomings we initiated the present study.

Hypotheses:

1. The EEG changes observed during HV and measured by the modified HV-Score correllate to an impaired CNS function.

2. Transcutaneous measurement of the PCO_2 will be easier to administer and tolerated better by the patient than the conventional analysis of the exhaled gas.

Methods and investigation:

In order to investigate the two hypotheses, 59 randomly selected pilots and candidates were included in the study.
The 59 persons underwent the EEG procedure with the addition of surface EMG recording from the dominant forearm and two recordings of CO_2, allowing to standardize the HV (and compare the methods to register the CO_2) and to do a simple reaction time test.
Standardizing the HV:
In order to investigate HV, we had to standardize it. Since the relevant changes under HV seem to result from hypocapnia we excluded simple measuring of respiratory volume or O_2-uptake and concentrated on alveolar CO_2 (via exhaled gas, see below) and peripheral carboxaemia (via transdermal measurement, see below).
In the literature there are different values concerning the aim value for CO_2:
18+/- 2mm Hg CO_2 (Adler, 1991), 21mm Hg for 3 min (Kenealy et al, 1986), lowering the pCO_2 by 18 mmHg (Achenbach et al, 1994).
We decided to set the aim for expiratory CO_2 at 20 mm Hg, to be reached after one minute and to be held throughout the fourth minute of HV.

Measurement of CO_2:

1. The standard method is to measure CO_2 via expiratory gas analysis.We used an ELIZA+ analyzer (Engström, Sweden) which measures CO_2 with infrared absorption in a bypass sample of 100 ml/min.
2. The transcutaneous analysis was done with a TINA TCM3 monitor (Radiometer Copenhagen, Denmark) which uses a ceramic solid electrode to measure O_2 and CO_2.

Reaction time measurement:

Since psychomotoric testing is normally nearly impossible without massive artefacts in the EEG, a method had to be devised which would indicate the subject´s performance and its possible impairment through HV without disturbing the EEG.

To avoid artefacts, movements had to be kept at a minimum and any need to open the eyes or speak had to be avoided.

Therefore a simple reaction time test is devised:

The subject has to react to a flash of the stroboscope by flexing his fingers which is recorded by skin electrodes over the long finger flexing muscles (on the dominant arm).

Both flash and EMG-reaction are recorded on the EEG printout. (To allow easier measurement, the printing is done at 60mm/sec paperspeed.)

To eliminate training effects, training is done before the first registration.

EEG analysis:

To assess the EEG changes during hyperventilation, a visual analysis is carried out using the HV-Score (see above, fig.1).

This way it is possible to score the HV-induced changes that take place at the time the reaction time test is performed.

Results:

Participants:

59 persons (27 pilots and 32 candidates) of those participating had complete datasets and were included in the following study.

The group was all-male, the mean age was 26,2 years (SD=9y), they were randomly selected volunteers, all were healthy.

CO_2 Measurement:

The first method (gas analysis) is faster and more reliable. We varied the way to sample gas and tried to avoid the mask but found it impossible to accurately measure the exhaled gas without a conventional mask (to prevent room air from getting sucked in the turbulent exhalation stream and thus contaminate the exhalation). The analyzer did not require any specific maintainance, calibration is easy and seldom required. Generally the pilots were used to wearing masks and the candidates tolerated it well.

Since the CO_2 concentration in the exhalation gas directly resembles blood levels, this method quickly showed changes in the CO_2 concentrations after the beginning or end of HV. Typically the CO_2 levels fell to 25mm Hg in the first minute and reached 20 mm Hg in the second minute of HV.

The second method (transcutaneous measurement) reacted much slower. In the above mentioned typical case the transcutaneous CO_2 would lag about two minutes behind the respiratory values: Only during the fourth minute the aim value of 20 mm Hg would be reached.

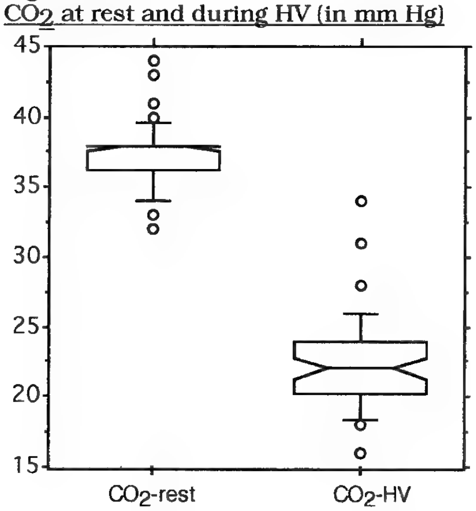
While the transcutaneous method did not require to bother the patient with a mask, it bothered the EEG-assistant with the need to regularly change the lining of the electrode, to constantly check and calibrate the machine. All in all the second method was kinder to the patient but more expensive, slower and more susceptible to artefacts and time consuming maintainance.

Though we used both methods on all reported 59 persons, only the readings of the gas analyzer are mentioned here.

Though many of the participants reached the aim of CO_2=20 mm Hg during HV, some had difficulties to maintain the necessary breathing volume or complained of dizziness or paresthesia. They were not forced to reach the aim of 20 mm Hg, but encouraged to keep on hyperventilating.

Figure 2:

CO_2 at rest and during HV (in mm Hg)

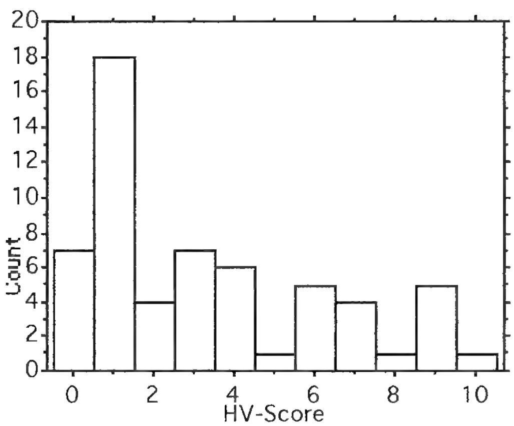


For all 59 participants the mean partial pressure for CO_2 during rest was 37,2 mm Hg (SD= 2,3) and during HV 22,1 mm Hg (SD=3,2). There was no correlation between CO_2 at rest and at HV.

EEG changes during HV:

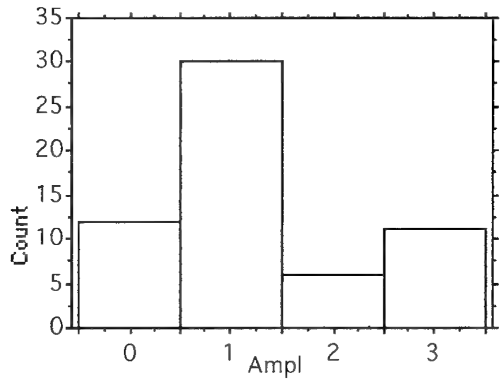
The EEG of the participants showed the usual buildup. The corresponding HV-score had a mean of 3,3 points (SD=3). The distribution of HV-Scores among the participants is shown in figure 3.

Fig. 3:
HV-Score (total)



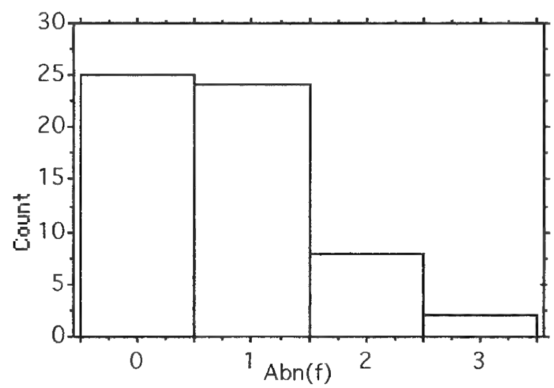
No one exceeded the limit of 10 points.
Differentiating the HV-score towards the changes during HV, one can observe alterations of the amplitude and changes of the alpha band main frequency and different normalization times. We did not observe focal abnormalities or paroxysmal discharges.
An increase of the amplitude of the EEG activity, however, was observed regularly:

Fig. 4:
Changes of the amplitude (0 points= the same amplitude under HV as at rest, 1 point= up to 1.5 times increase in amplitude, 2 points= up to doubled amplitude, 3 points= more than doubled amplitude under HV)



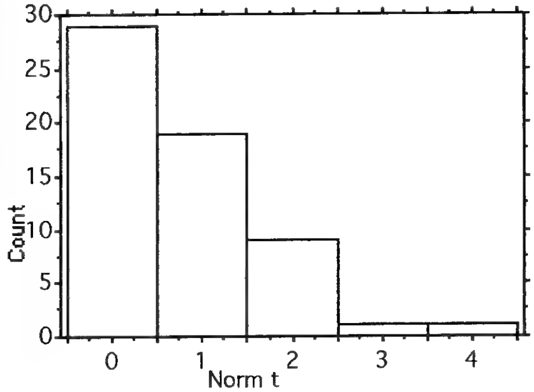
A slowing of the EEG activity was also observed often:

Fig. 5:
Slowing by more than 2 Hz under HV (1 point= after more than 2 min of HV, 2 points= after 1-2 min, 3 points= after less than 1 min)



It is seen that while more than half of our subjects exhibited slowing of their EEG activity, most of them do so only after two minutes of HV.
The normalization of the EEG after HV is shown in figure 6.

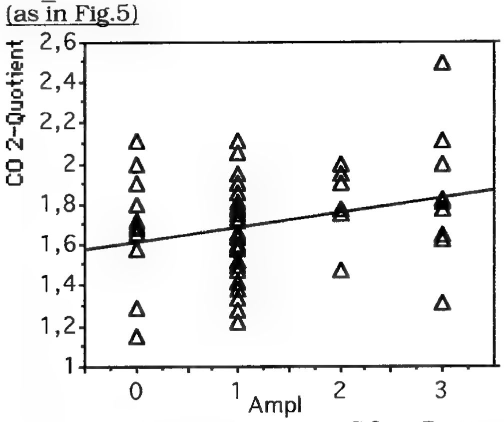
Fig. 6:
Normalisation of EEG after HV (0 points= during HV, 1 point= less than 30 sec, 2 points= less than 1 min, 3 points= less than 2 min, 4 points= more than 2 min after end of HV)



While the EEG activity of most participants normalizes after 30 seconds, only two take longer than a minute.

Correlations between HV-Score and CO_2:
The effect of HV on CO_2 levels is shown best by dividing the level of CO_2 at rest through the level of CO_2 during HV (CO_2-Ratio)
Even if the expected influence of HV on HV-Score was observed, no significant overall correlation could be shown between CO_2 and HV-Score. However, there was a significant (if weak) correlation between factors of the HV-Score and a Ratio of CO_2 at rest and CO_2 during HV.

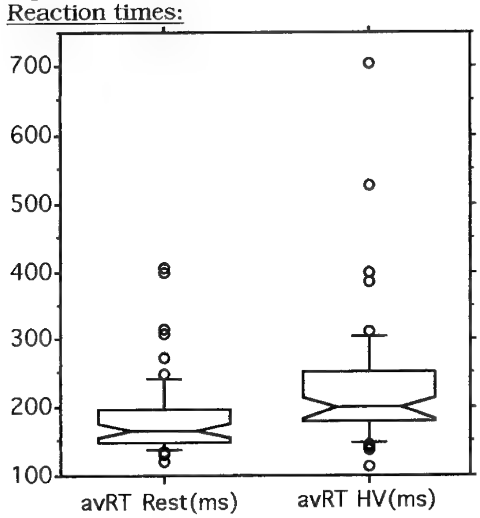
Fig. 7:
CO_2 Ratio (at rest/HV) and change of amplitude (as in Fig.5)



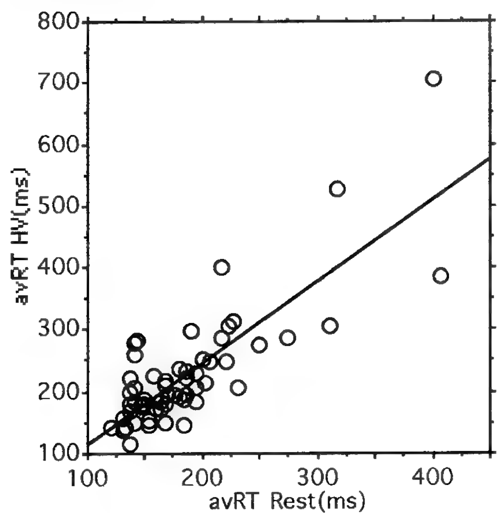
The correlation between CO_2 Ratio and amplitude change of the EEG activity reached the level of significance ($r=.274$ and $p= .037$): The stronger HV is, the more changes in the EEG activity it produces.

Reaction Time during HV:
As was to be expected, reaction time changed under HV.

Fig. 8:
Reaction times:



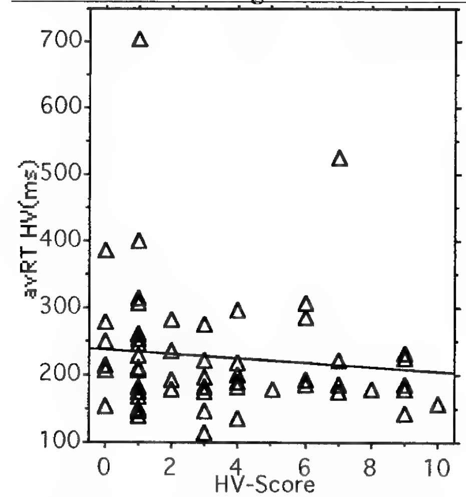
Average reaction time at rest was 184 ms while it was 226 ms during HV.
This marked slowing showed the expected effect of cerebral hypoxia that results from HV.
Persons reacting slower at rest usually took longer under HV as well (correlation $r=.81$, $p<.0001$).
Fig. 8 b:
Correlation between reaction times at rest and during HV:



Although reaction times were doubtless affected by HV, the variance of the reaction times could not be explained by CO_2 or changes of EEG activity. (The found reaction times or the ratio reaction time at rest/during HV could not be

correlated to either the level of CO_2 or the HV score of the EEG.)

Fig. 9:
Reaction times during HV and HV score:



Since the correlation between HV score and reaction time during HV is too weak (r=.1) to be interpreted, one can only say that at least the hypothesized impairment of cognitive functions is not detectable by our experiment (on the contrary, the persons with more changes in their EEG activity seem to react faster under HV).

Further analysis could not reveal meaningful correlations between HV-Score and
-reaction time in our experiment or
-the reaction time measurements done in more complex tests (ERT) at the psychological department or
-the result (grading) of the total investigation at the psychological department of the GEAF IAM.
Our measurement of reaction time ratio corellated with the ERT results (r=.42 ; p=.017) but not with other variables and an analysis of variance failed to detect meaningful connections.

All in all, the data seemed to be consistent and showed changes of EEG activity and reaction times under HV.
The data could not, however, substantiate Hypothesis 1, because there was no correlation between reaction time and the HV-Score.

Discussion:

Hypothesis 1. (The EEG changes observed during HV and measured by the modified HV-Score correllate to an impaired CNS function.):
Conventional wisdom held that the observed changes of the EEG activity during HV resembled changes in the cognitive processes.
To investigate this connection, the revised HV-Score (fig. 1) and a reaction time measurement were used. Although the latter is a rather simple instrument to measure the cognitive state, it has the advantage of being useable during the recording of the EEG. The disadvantage of our reaction time measurement is its simplicity: It does not include complex decision making and may thus be only a weak representation of overall cognitive functions or intelligence.
We know, however, that there are other methods that are better suited to investigate the consequences of HV on the cognitive state. But our aim was different:
We wanted to investigate wether it is useful to select candidates for military flying duties on the criterion of the stability of their EEG and hope that the ones with the stabelest EEG are the ones most resistant to stress of all kinds,
or wether it is better to just exclude the candidates at risk to develop epileptic seizures and do no futher interpretation of the EEG.
Since we are dealing with a situation in which candidates are selected and today are speaking about ways to enhance situational awareness, it is important to think about the usefulness of our selection procedures towards the aim of selecting the candidates with the potential to reach the best situational awareness (without being disrupted by instable resources).
Here we have to choose between two concepts: Agility vs. stability.
The reason for selecting towards agility is simple: Faster and more intelligent candidates have more cognitive resources and can adapt more easily to new situations. By selecting this way, the conditions for achieving good situational awareness are met.
The reason for selecting towards EEG stability is also understandable: Maximum instability of the EEG is found in epilepsy, and aviators with EEG instability but without manifest epilepsy have been found to be four times more likely to cause crashes than normal pilots (Lennox-Buchthal et al, 1959, citation in Trojaborg, 1992). Also, instability of the EEG may resemble an instability of cognitive functions that would hinder situational awareness under stress.
Since stability of the EEG is acquired by a dynamic process of constant inhibition, it has its prize by inhibition of all processes, including desired speed.
The strategy implemented so far at the GEAF IAM was to promote agility by choosing the candidates with the best results in the psychomotoric tests of the psychologic department of the GEAF IAM and then sort out the ones with the most unstable EEG reaction in the medical department of the GEAF IAM.
This had the disadvantage of later eliminating often the most promising candidates (rated `A`) because of their greater instability of the EEG.

The results of the current investigation do not substantiate the theory of the indication of instable cognitive functions under changing EEG activities, so that the first hypothesis has to be rejected.
As a consequence, one could state that the electrophysiological phenomena that are recorded in the EEG should be treated as just that. Only signs of pathological excitability should be evaluated, and there should be no interpretation made concerning the possible instability of cognitive functions.
Since there seems to be no further slowing of the reaction time in subjects with more visible EEG changes, it does not make sense to be overly cautious regarding the EEG changes during HV.
During the investigations of this matter, the attitude at the GEAF IAM towards EEG changes during HV such as increase of the amplitude or gradual slowing of the EEG activity has grown more permissive, so that far fewer candidates are rejected now because of their EEG changes during HV.

Hypothesis 2. (Transcutaneous measurement of the PCO_2 will be easier to administer and better tolerated by the patient than the conventional analysis of the exhaled gas):
While the gas analysis followed the process of HV fast and closely (the onset of paresthesia and of EEG changes coinciding with the drop of CO_2 towards 20mm Hg), the transcutaneous measurement lagged behind around two minutes on average.
Also the mask was rather well tolerated (may be because of the aeromedical setting in which a mask seems rather natural).

The self adhesive fixation for the transcutaneous electrode was easily placed on the forearm and well tolerated, but the machine had to be calibrated after each measurement and produced more errors and consumed more precious time.

So the second hypothesis yields a mixed answer:
With neither method did we receive any complaints from our participants.
If the participants were to move around or to speak, the transcutaneous electrode would prove better. But under the circumstances of the EEG recording, the gas analysis via mask proved to be faster, more reliable and easier to use.

Conclusion:

1.:
With our method of evaluating EEG changes during standardized HV (HV-Score) and simultaneously measuring cognitive functions (reaction time), we cannot demonstrate a correlation between the two.
There should be made a strict differentiation between electrophysiological phenomena and their stability as observed in the EEG and evaluations of cognitive functions.

2.:
The measurement of CO_2 during EEG is useful for standardizing HV. The method of analyzing breathing gas sampled from a mask proved to be faster, easier to use and more reliable while being tolerated as well as the transcutaneous electrode measurement.

References:

Achenbach-Ng J; Siao TC; Mavroudakis N; Chiappa KH; Kiers L:
Effects of routine hyperventilation on PCO_2 and PO_2 in normal subjects: implications for EEG interpretations.
J Clin Neurophysiol 1994 Mar;11(2):220-5.

Adler G: Hyperventilation as a Model for Acute Ischaemic Hypoxia of the Brain: Effects on Cortical Auditory Evoked Potentials.
Eur Arch Psychiatry Clin Neurosci (1991) 240: 376-9.

Fowler B, Taylor M: The effects of hypoxia on reaction time and movement time components of a perceptual-motor task.
Ergonomics, 1987, Vol 30, No 10, 1475-1485.

Fowler B; Kelso B:
The effects of hypoxia on components of the human event-related potential and relationship to reaction time.
Aviat Space Environ Med 1992 Jun;63(6):510-6

Fowler B; Lindeis AE:
The effects of hypoxia on auditory reaction time and P300 latency.
Aviat Space Environ Med 1992 Nov;63(11):976-81

Fowler B, Prlic H, Brabant M: Acute Hypoxia Fails to Influence Two Aspects of Short-Term Memory: Implications for the Source of Cognitive Deficits.
Aviat Space Environ Med 1994; 65: 641-5.

Freund W:
The Neurophysiological Methods Used to evaluate the Fitness for Military Flying in the GAF Institute of Aviation Medicine.
XXX th International Congress on Military Medicine Augsburg, June 5-11, 1994.

Glaser H-J and Freund W:
Neurological Decision-Making in Aviation Medicine Based on Electrophysiological Methods.
AGARD Conference Proceedings 553 (ISBN 92-836-0003-7), 1994, pp 19/1 ff.

Hoffelt W and Gress W: Validation of Gradings at the German Air Force Institute of Aviation Medicine´s Department of Psychology. Unpublished manuscript, 1992.

Jensen AR; Whang PA: Reaction times and intelligence: a comparison of Chinese-American and Anglo-American children.
J Biosoc Sci 1993 Jul;25(3):397-410

Kenealy J A, Penovich P E, Moore-Nease S E: EEG and Spectral Analysis in Acute Hyperventilation.
Electroenceph clin Neurophysiol, 63 (1986) 98-106.

Konishi T:
The standardization of hyperventilation on EEG recording in childhood. I. The quantity of hyperventilation activation.
Brain Dev 1987;9(1):16-20

Kraier V, Van Huffelen A C, Wieneke G H:
Changes in quantitative EEG and blood flow velocity due to standardized hyperventilation; a model of transient ischaemia in young human subjects.
Electroenceph clin Neurophysiol, 70(1988) 377-387.

Kraier V, Van Huffelen A C, Wieneke G H, Van der Worp H B, Bär P R:
Quantitative EEG changes due to cerebral vasoconstriction. Indomethacin versus hyperventilation-induced reduction in cerebral blood flow in normal subjects.
Electroenceph clin Neurophysiol, 82(1992) 208-212.

Lindenberger U; Mayr U; Kliegl R:
Speed and intelligence in old age.
Psychol Aging 1993 Jun;8(2):207-20

Link CG; Leigh TJ; Dennison JK:
The effects of BRL 46470A, a novel 5-HT3 receptor antagonist, and lorazepam on psychometric performance and the EEG.
Br J Clin Pharmacol 1993 Apr;35(4):395-9

Myung JA; Lynn R:
Reaction times and intelligence in Korean children.
J Psychol 1992 Jul;126(4):421-8

Navathe P D and Singh B: An Operatinal Definition for Spatial Disorientation.
Aviat. Space Environ. Med. 1994; 65: 1153-5

Njemanze P C, Antol P J, Lundgren C E G:
Perfusion of the Visual Cortex During Pressure Breathing at Different High-G Stress Profiles.
Aviat Space Environ Med 1993; 6 4: 396-400

Paul M A, Fraser W D: Performance During Mild Acute Hypoxia.
Aviat Space Environ Med 1994; 65: 891-9

Sieber FE; Derrer SA; Eleff SM; Koehler RC; Traystman RJ:
Hypocapnic-hypoglycemic interactions on cerebral high-energy phosphates and pH in dogs.
Am J Physiol 1992 Dec;263(6 Pt 2):H1864-71

Trojaborg W: EEG Abnormalities in 5,893 Jet Pilot Applicants Registered in a 20-Year Period.
Clinical Electroencephalography 1992, Vol 23 No. 2, p72-8.

Van der Worp H B, Kraier V, Wieneke G H, Van Huffelen A C:
Quantitative EEG during progressive hypocarbia and hypoxia. Hyperventilation-induced EEG changes reconsidered.
Electroenceph clin Neurophysiol, 79(1991) 335-341.

Velho-Groneberg P:
Metabolische Enzephalopathien. (lecture)
29. EEG- Fortbildungskurs München, 27.02.-04.03.1995
Munich, Germany.

Vidulich M A, Stratton M, Crabtree M, Wilson G:
Performance-Based and Physiological Measures of Situational Awareness.
Aviat Space Environ Med 1994; 65 (5, Suppl.) : A7-12

Yamaguchi F, Meyer J S, Sakai F and Yamamoto M:
Normal human aging and cerebral vasoconstrictive responses to hypocapnia.
J Neurol Sci, 1979, 44:87-94.

Author´s adress:

OSA Dr. Wolfgang Freund
Fachgruppe Neurologie/Psychiatrie
Abteilung I
Flugmedizinisches Institut der Luftwaffe
Marseillestraße
82242 Fürstenfrelbruck
Germany

PHYSIOLOGICAL VESTIBULAR LIMITATIONS OF MOTION PERCEPTION IN AVIATION ENVIRONMENT

I. I. DIAMANTOPOULOS, MD, Leut (HAF)

Otolaryngology Department , 251 Air Force Hospital
115 25 Athens, Greece .

SUMMARY

Perception of angular & linear acceleration provides together with vision the fundamental cues for situational awareness in aviation environment.

Therefore, the function of the otoliths & the semicircular canals as the end-organ of balance and their perception limitations play an important role in a flight environment employing mainly high performance aircraft.

This paper addresses the need to identify & provide the physiological basis for the vestibular limitations of motion perception which accounts for several disorientation illusions. Under this view, certain areas of flight configuration envelope known as precipitators of illusions are outlined and the consequent most common vestibular illusions are attributed to their physiological basis.Vestibular behaviour continuoum is also outlined both in a non-1G environment and in Motion Sickness.

The danger upon flight safety imposed by the veestibular function limitations can be alleviated by proper training, selection & noumerous technology aids integrated in modern cockpits.

INTRODUCTION

Spatial Orientation can be defined as the pilot's accurate perception of position, motion or attitude of him/her self or of the aircraft with respect to the gravitational vertical of the Earth.

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

On the other side, **Situational Awareness** derives from the correct realization of events in the outside world together with the status of oneself within.
Profoundly, vestibular physiology holds an important contributing role on the above by means of limitations of vestibular function, based on structural (anatomic) & dynamic (physiological) elements, described as follows .

1. LABYRINTH ANATOMY

The sense organs embedded bilaterally in the petrous part of the temporal bone of man are called the labyrinth organs, or

collectively the labyrinths.

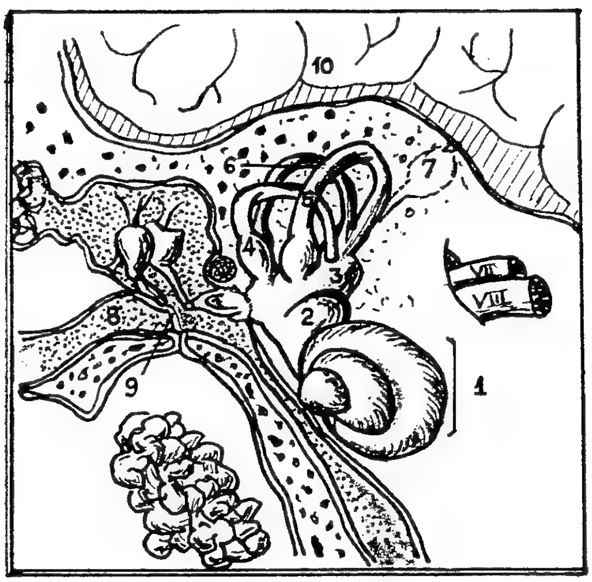


FIG 1 Schematic representation of the Labyrinth organs. 1:Cochlea, 2:Saccule, 3:Utricle, 4,5 & 6:Semicircular Canals (4:Horizontal, 5:Ant.Vertical,6:Post.Vertical), 7:Endolymphatic Sack, 8:Ext.Auditory Meatus, 9:Tympanic Membrane, 10:Cer.Temporal Lobe,

They include (FIG 1)

(a) an anterior spiral part sensitive to acoustic vibrations, the cochlea, and

(b) a posterior part the vestibule, consisting of

(1) The **Saccule** and the **Utricle**, two sack-shaped structures sensitive to linear a cceleration imposed upon the head/body and

(2) The three **semicircular canals** cyclically constructed, responsible for perception of angular head accelerations at any plane possible.

Ducts interconnect anterior and posterior labyrinth and the saccule - canals complex to the saccule. The whole system is filled with fluid close to water in density, called the endolymph while the actual temporal bony cavities containing the labyrinths are filled with the perilymph.

The endolymphatic duct connects the endolymph containing structures with the endolymph reservoir called the (FIG 1) **endolymphatic sack**, while there is a communication between Cerebrospinal fluid (CSF) and the labyrinth through the cochlear aquaduct.

2. THE OTOLITH ORGANS

2.1 Otolith Structure & Dynamics

The otolith organs include the saccule and the utricle, both sack-shaped, located within the posterior labyrinth.
Utricle is in conjunction with the semicircular canals lumen while the saccule is located beneath, at about right angles to each other (FIG 2).

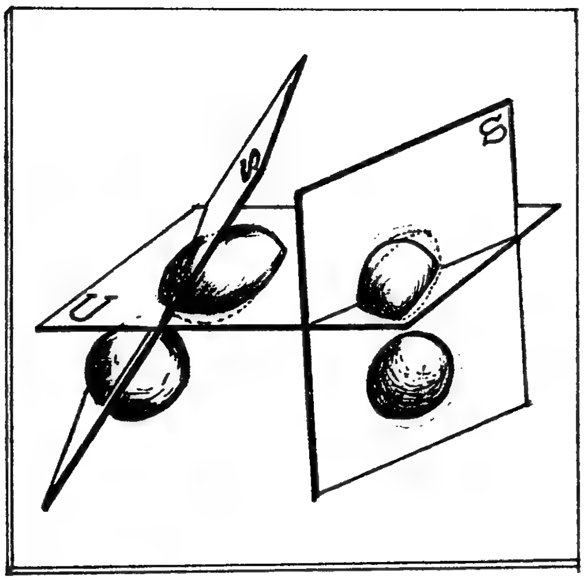


FIG 2: Saccular - Utricular level relationshipin space U:utr. level, S: saccular level

They both contain a sensory epithelium called **the macula**[33,55]. The macula surface is covered with an epithelium consisting of supporting cells and ciliated cells bearing 60-100 stereocilia of different size and a kinocilium(FIG 3).
Stereocilia are graded from the smaller to the larger in length, which lies next to the kinocilium, the largest cell cilium.

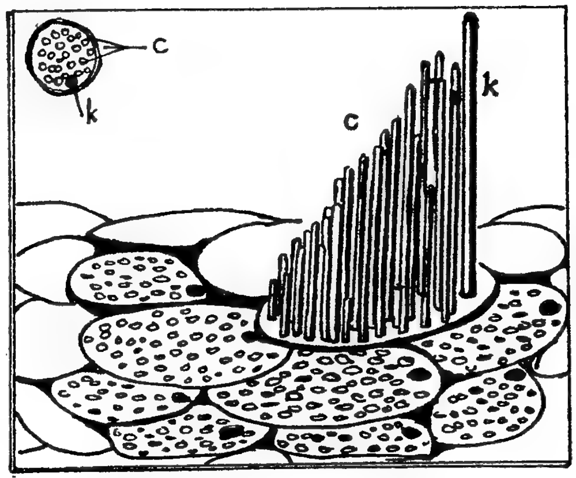


FIG 3: Schematic Representation of the maculae. H:Hair Cells, C: Stereocilia, K:Kinocilium

There are two types of hair cells: (FIG 4) The flask shaped *Type I cell*,enclosed alone or with others into a nerve chalice deriving from large diameter afferent nerve fibre, and the cylindrical *type II cells* feeding into a small diameter nerve fibre.

All cilia protrude into a gelatinous substance, the **otolithic membrane,** containing on its outer surface a mass of

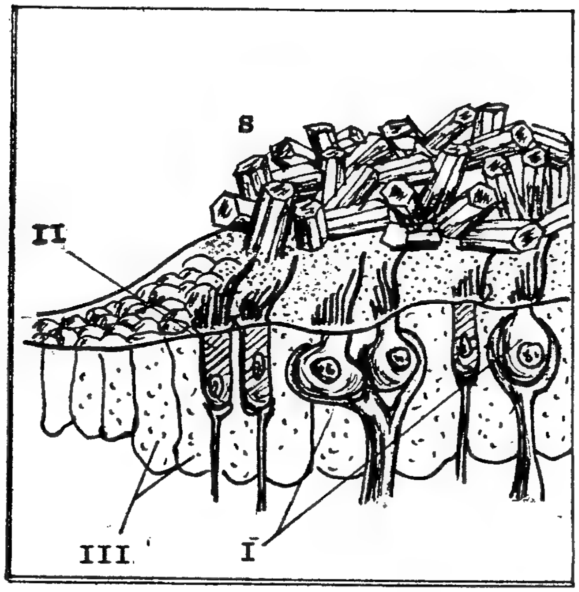


FIG 4:Otolithic Organ strycture diagramm I:Hair Cells type I. II: Hair Cells Type II, III:Supporting Cells, S:Stratoconia

calcite crystals, known as **otoliths** or **stratoconia.**
Otoliths bear a density of 2.74[10], almost three times that of the surrounding endolymph and occupy a plane of 1.5 - 2 mm^2 kept in place by elastic filaments attached to the organ walls.

Because of this density difference, the sensory hair are bent as the attitude of the head changes with reference to the gravitational vertical, therefore generating the signal for head position relative to the earth surface.

The utricular maculae lie approximately in the horizontal plane, while the saccular ones occupy a vertical (sagittal) one, almost parallel to each other[2,3].
This arrangement conforms with the need to detect and identify direction of head motion or attitude of any linear acceleration imposed upon the head (FIG 2).
In this way, the rate of discharge of an afferent neuron signalling tilt of the head is a sin function of the angle of tilt[16,34].
Collectively, any linear acceleration acting on the body/head is seen by a certain angle which the acceleration direction vector forms with the horizontal (utricular) level and the vertical (saccular) level.

Type I cells demonstrate a regular afferent activity pattern corresponding to the displacement of the otolithic membrane (regular cells - tonic receptors) while *type II cells* fire irregularly and their afferent activity signals dynamic characteristics of the macular shearing force i.e. magnitude, rate of displacement (irregular cells - phasic receptors).

Afferents from the otolith organs project (FIG 5) to the ipsilateral vestibular nuclei in the brainstem : lateral (Deiter's) and medial nuclei [18]. Second or higher order neurons arising from the vestibular nuclei convey information to the cerebellum [40,31], the reticular formation [65] and the motorneurons of the neck to mediate postural reflexes, and to the thalamus & cerebral cortex.
Fibres to the oculomotor nuclei mediate ocular reflexes

associated with head tilt and linear acceleration [1].

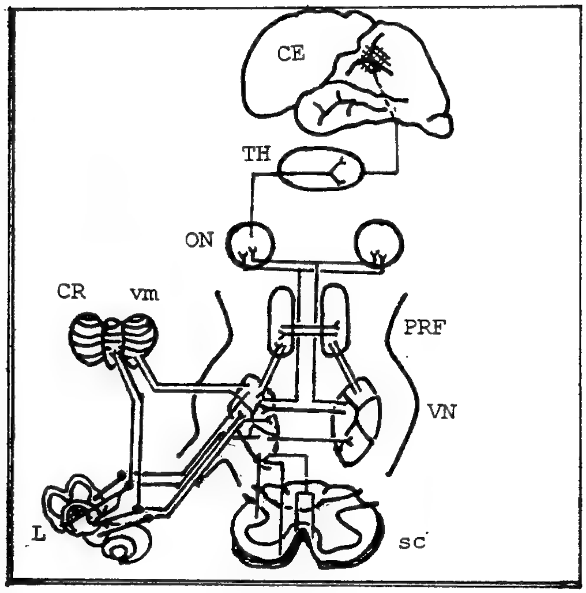


FIG 5:Labyrinth Organs afferents schematic representation.VN:Vestib.Nuclei,CR:Cerebellum(vm:cer.vermis),PRF:Pontine Reticular Formation, ON:Oculomotor Nuclei TH:Thalamus (ventro-post.areas) CE:Cerebral Reception Area

2.2 Linear Acceleration Thresholds

There are two stimulus situations[28] used for the identification of linear acceleration thresholds :

(a) Subjects are exposed to an oscillating accelerative force produced by moving them to & fro a fixed linear path, or to a sinusoidal stimuli by the use of the parallel swing and

(b) Subjects are exposed to a constant linear accelerative force that differs in direction or magnitude or in both with respect to gravity.

For the later situation , three methods are utilized:

(b1) Tilting the subject through a specific angle, so that the direction of the body tilt specifies the direction of the linear acceleration vector and the sine of tilt angle represents the magnitude of the shearing force.

(b2) Rotating the subjects in the centrifuge, and

(b3) Accelerating the subjects along a smooth linear track in a defined direction with respect to the body's long axis & the gravitational vertical.

The most popular of the above methods has been body oscillation along different axes while step acceleration and single acceleration sinusoid have been employed lately (TABLE I), bearing the following results .

Collectively, threshold values range at 0.3 Hz from 1.4 to 18 $cm.sec^{-2}$ for Z-axis.

Research on linear acceleration thresholds relates thresholds as a function of (a) acceleration axis of the body and (b) oscillation frequency spectrum.

In particular, thresholds demonstrate higher values for the detection of motion in the Z (longitudinal) axis of the body.

Moreover, threshold values for the X, Y axes(FIG 6) are probably 50 % lower than the ones for Z axis [7,2].

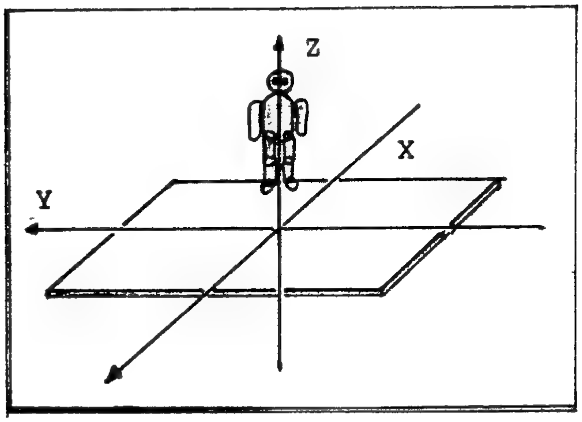


FIG 6: NATO Standard X,Y,Z Axes Terminology

The reason for this disparity between objective and subjective measures of otolithic activity may be that CNS employs a lower "gain" in the integration and processing of saccular than utricular afferents, since both are electrophysiologically equivalent , but utricular information carry more perceptive "weight" because they are more significant in everyday locomotion[7].

The appearance of threshold as a function of frequency with the otolithic sensitivity increasing over frequency (from 0.18 at 0.2 Hz to 0.05 $m.s^{-2}$ at 2.0 Hz)in the bandwidth between 0 - 2 Hz may be attributed to the primary contribution of type II hair cells (irregular units or phasic detectors) into signalling changes in linear acceleration , as described electrophysiologically by Fernandez & Goldberg[15].

When subjects are exposed to a sustained linear acceleration of constant magnitude the time taken (**t**) to detect stimulus varies with acceleration (**a**) , so that the product **a.t** is constant and carries the dimension of velocity[67] with an approximated value of 0.3 - 0.4 m/s.

3. THE SEMICIRCULAR CANALS

3.1 Structure & Dynamics

There are three semicircular canals adjacent to the utricle (**FIG 7**) at right angles between them , called the **horizontal**, the **posterior vertical** or inferior and the **anterior vertical** or superior canal.

The cavity or lumen of each elliptical in shape with a mean diameter of 0.3 mm and the anterior vertical canal larger[12] in radius (2.2 mm) of curvature than the others (1.6 - 1.9 mm).

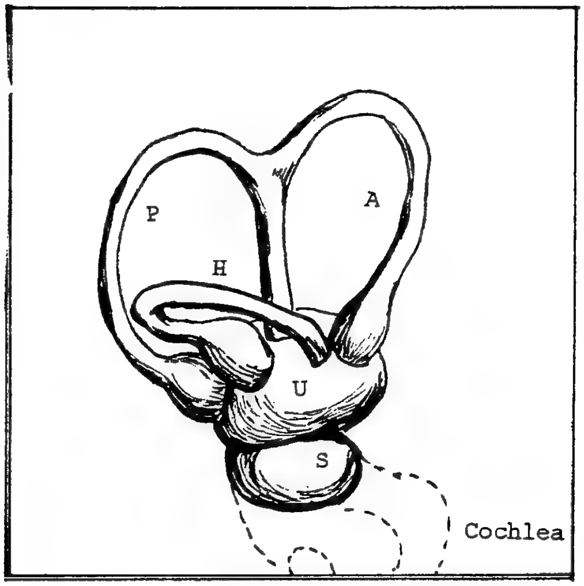


FIG 7: Canals'Diagramm. A:Anterior Vert., P:Posterior Vert., H:Horizontal Canal.

The horizontal canal lies in the horizontal plane of the body when the head is inclined about 25° forward with respect to the Horsley - Clark stereotaxic plane (plane through the auditory canals and the lower orbital margins).

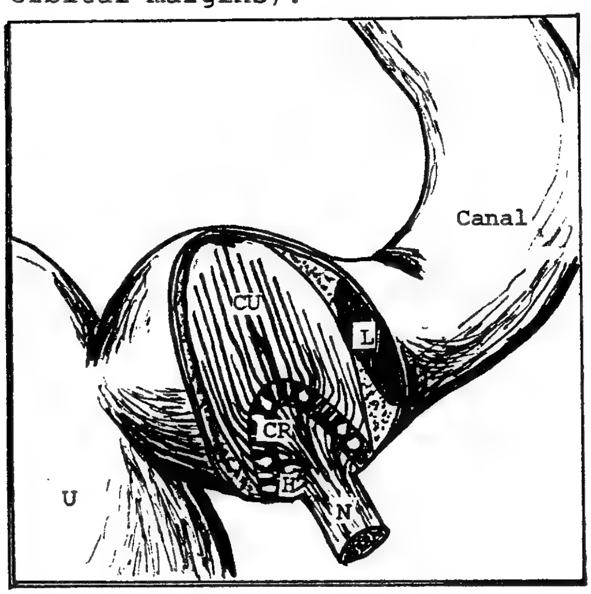


FIG 8: Canal Ampulla Representation. CR:Cristae, H:Hair cells, CU:Cupula, L:Lumen, N:NerveFibres, U:Utricle,

Each one interconnects with the utricle after forming a swelling, known as **the ampulla,** the housing of the neurosensory cells of the canals , which congregate to form a ridge, ***the cristae*** **(FIG 8).**

Their synthesis reflects the neurosensory epithelium of the otolithic organs, consisting of supporting cells and hair cells of type I (regular units, tonic detectors) and type II (irregular units, phasic detectors).

The hair like cell projections, or cilia are covered by a gelatinous structure, the cupula, forming a water-tight cross section diaphragm across the canal.

Angular head movements are followed by the canal structure, while endolymph inertia causes deflection of the cupula which in turn through cilia bending will alter the signal of the canal, thus producing the cue for sensing angular turn[60].

Source	Stimulus	f (Hz)	Threshold ($cm.sec^{-2}$) X	Y	Z	No Subjects
Mach (1875)	Oscillation	0.14	@	@	10-12	1
Travis & Dodge (1928)	Oscillation	0.12-0.2	8.0	5.0	@	2
Walch (1962)	Oscillation	0.3	@	@	8.2 (5.9)	13
Walch (1964)	Oscillation	0.25	5.5	@	@	7
Parker et al (1978)	Oscillation (prone)	0.37	@	@	6-12	6
Walch (1961)	Parallel swing	0.4	@	3.8	5.3	6
Greven et al (1974)	Parallel swing	0.29	4.5	3.5	5.3	12
Benson et al (1986)	Continuous oscillation	0.3	2.5	3.2	7.0	6
Meiry (1966)	Step acceleration		5.9	@	9.8	3
Benson et al (1987)	Acceleration sinusoid	0.3	6.3	5.7	15.4	24

TABLE I : Linear Acceleration Thresholds

The canal which is closer to the plane of rotation will be maximally stimulated, but a random rotation at any plane will stimulate all three canals according to the proximity of rotation level to their own.

The three canals of the right labyrinth form functional pairs <u>synergists)</u> (**FIG 9**) with the similar ones on the opposite site, so that a horizontal turn towards i.e. the right, will depolarize right horizontal canal hair cells and hyperpolarize the opposite ones, thus increasing the ipsilateral to the turn afferent signal and decreasing the contralateral one.

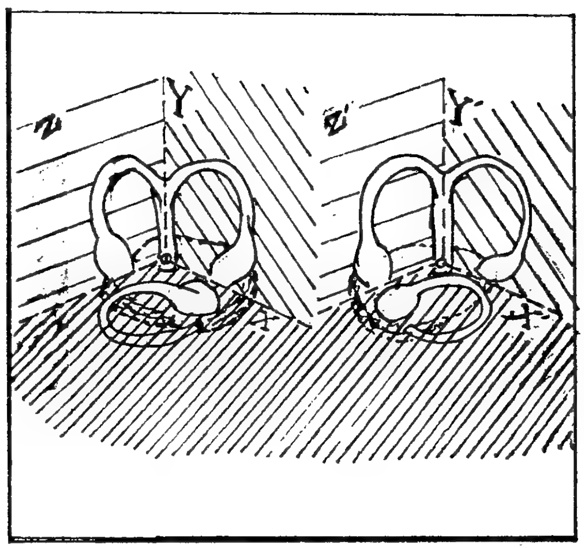


FIG 9: Canal Synergistic Pairs

In general, (**FIG 10**) any rotation provoking displacement of the cupula towards the utricle *(ampullofugally)* will increase afferent signal, while

cupula bending away from the utricle *(ampullopetally)* will decrease canal signal [39].

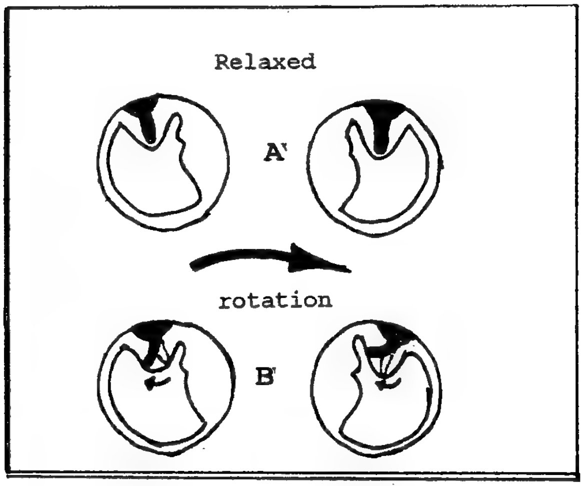


FIG 10: **The Cupula rotation effect: A:Canal Pair Relaxed, B:during rotation to the right.**

Afferents from the canals drive the signals to the brainstem vestibular nuclei (superior & medial).
At this level, activity of different types of cells (I, II, III, IV) constitutes the physiological basis for **the push - pull model**[28] responsible for transdusing the differential canal signal **(FIG 5)** through the longitudinal fasciculus and pontine & mesencephalic reticular formation to the oculomotor nuclei[13], the cerebellum, thalamus and cerebral cortex areas 2v &3a[17].

3.2 Angular Acceleration Threshold

Three manifestations of vestibular - canal activity have been used[28] to determine threshold values:

1.Report of feeling of rotation

2.Nystagmus

and **3**.The oculogyral effect (that is the apparent displacement of a point of light in the dark, provoked by vestibular stimulation).
Considerable diversity of the threshold measures exist, partly justified by the variety of the psychophysical methods of investigation:

Method of Limits: includes the report of the first motion sensation with increasing stimulus magnitude and yields a $0.2°s^{-2}$ [61] or $0.25\text{-}0.7°s^{-2}$ [27] threshold value.

Double Staircase Method : Supra & subthreshold stimuli alternate until a point of convergence which represents threshold [57] of $0.05\text{-}2.2°s^{-2}$

Signal Detection Procedures: Subjects are presented with stimuli and the responce latency

Study	Subj.	Stimulus	Responce	Thresh. Acceler. ($^{\circ}/s^2$) X	Y	Z
Groen &Jonkees (1948)	30	Rotating chair	First mot. perception		0.23-2.0 (mean=0.8)	
Clark & Stewart (1972b)	92	Rotating chair	First mot. perception		0.05-3.18 (m=0.44)	
Benson,Hutt (1989)	30	Rotating chair	First mot. perception		m=1.5	
	20	Rotating litter		m=2.04	m=2.07	
Graybiel et al (1948)	5	Human Centrifuge	Oculogyral illusion		m=0.12	
Clark & Stewart (1972b)	92	Rotating chair	Oculogyral illusion		0.03-0.59 (m=0.11)	
Miller & Graybiel(1975)	300	Rotating chair	Oculogyral illusion		0.02-0.95 (m=0.10)	

TABLE II : Angular Acceleration Thresholds

of stimuli detection is correlated to stimulus magnitude Threshold Value at $0.3^{\circ}s^{-2}$..
Cupulometry: threshold derives from angular velocity that just fails to produce postrotatory responses. Calculated value of $0.25 - 0.45^{\circ}s^{-2}$[23].

TABLE 2 provides threshold estimates in correlation to each other, because investigation & study methodology allow comparison.

As anticipated, when subjects are presented with a visual cue, thresholds lower because they are provided with additional motion cues by virtue of the oculogyral illusion.

By far,for perception of angular movements of short duration (t<= 5 sec) the important parameter is angular velocity, which must exceed 0.2-8.0o/s (mean 1.5° /s) while in sustained rotation (t>=10 sec) angular acceleration should exceed mean value of 0.3o /sec2 ($0.05-2.2^{\circ}s^{-2}$)[23].

4. VESTIBULAR BASIS OF COMMON DISORIENTATION ILLUSIONS

Functional and dynamic characteristics of the vestibular organs limit the ability of non-visual sensory systems to respond correctly (detect or estimate) the attitude and motion of the aircraft, therefore imposing a potential hazard upon flight safety.

4.1 Motion Threshold disorientation:The Leans

The erroneous perception of roll attitude[2] within flight in known as **"the leans"**.

It is a common manoeuvre for the pilot to enter a turn smoothly and gradually, keeping the turn's angular velocity at subthreshold value ($0.2-8.0^\circ\ s^{-2}$).

Meantime, otoliths and gravireceptors fail to signal change, since in such a coordinated turn the apparent vertical is aligned with the long (Z) body axis .

Thus, a suprathreshold recovery will provoke the erroneous perception of flying with one wing low, when the aircraft horizon suggests a level flight.

The mismatch will spontaneously resolve as soon as unambiguous external visual cues appear.If not , the pilot in an effort to accommodate both cues, aligns his body with the perceived vertical and flies level.

On occasions, a different threshold between left & right side canals may exist.A turbulent flight [2] inducing oscillatory motion to the aircraft may lead to motion detection on only one side, rending again the pilot disorientated with the leans.

4.2 Canal Misperception of Angular Motion

As stated earlier, canals act for short duration angular motion as <u>angular speedometers</u> and for sustained as <u>angular accelerometers.</u>

In this way, after exposure to a sustained (t>30") and steady angular motion(i.e. prolonged spin, coordinated turn) the canal cupula will return to pre-displacement status, with the rotation signal fading away.

The time taken for signal amelioration varies, apart from cupula deflection characteristics ,as a function of rotation speed and axis, complementary visual & gravireceprors' cues and pilot's habituation to the flight configuration [42].

4.2.1 Canals' Somatogyral Effect

When the pilot has allowed sufficient time for the rotation signal to subside,recovery from the manoeuvre will provoke illusory rotation signal towards the opposite side, and on the same rotation plane.

It is the endolymph inertia which will deflect cupula oppositely, with the pilot's view degraded by the resulting nystagmus until adequately suppressed.

This perception limitation of the canals known as somatogyral illusion endangers flight safety since the inexperienced pilot in an effort to alleviate the illusion may put the aircraft back into rotation.

The consequent flow of illusory information will render aircraft recovery unlikely (*graveyard Spin*).

Retinal image for postrotatory nystagmus smear is greater in yaw (horizontal) plane than in other canal planes.So,nystagmus suppression and recovery is longer in this plane.

4.2.2 Canals' Coriolis Effect

Once a prolonged angular motion has allowed cupula to completely or partly return to its resting position, a head movement in another plane of rotation will bring about illusory rotatory sensations, known as Coriolis Illusions, deriving from the co-effect (cross-coupled) of simultaneous rotation in two planes.

The principal effect of head movement in one plane during rotation in another, results in an illusory perception of rotation in the third orthogonal plane .

Aetiology for Cross-coupled illusion lies in the change in angle from which each canal sees the new plane of rotation .

Oculogyral component and increased intensity of the effect have been reported[26] when the head moves out of rotation level during deceleration from a sustained turn, as in the case of an instrument landing descent where the head turns to select a new R/T frequency.

4.3 Otolithic Misperception of Linear Motion

The acceleration of gravity is man's physical environment and possesses two basic characteristics:

1. It is constant and vertical

&2. Can not be distinguished from any other linear acceleration acting on the body --Stott's [56]third rule- by means of otolith organs.

4.3.1 Otoliths' somatogravic effect

Therefore, any sustained linear acceleration will melt with gravity and the resultant vector will be the new vertical, because of high expectancy that a sustained acceleration is by assumption gravity, hence vertical[66], providing with illusory attitude cues, the somtogravic illusion.
Consequently in a flat turn where the axis of the resultant vector is not aligned with the body's Z axis, the pilot feels a banked-out attitude when flying wing-level.

Accordingly, pitch attitude errors are associated with changes in the longitudinal aircraft acceleration[21].
Forward acceleration with gravity causes a backwards rotation of the resultant vector, which being taken as vertical , induces a pitch-up illusory attitude.

In a descent with no external reference, the increase in aircraft speed can provide the pilot with a pitch- up attitude cancelled by the true pitch- down attitude of the descending aircraft.No recovery action may be undertaken in a plane loosing height rapidly[11].

4.3.2 Atypical Otolith Stimulation (G-Excess)

During a sustained large radius 2G turn with a liminal turn rate (4°/s) angular head motion induces illusory perception of aircraft attitude [19], other than Coriolis, attributed to otolith responce to abnormal (*2G*) transient stimulation following changes in the abnormal force vector [2] orientation.

5. OCULAR COMPENSATION FOR MOTION PERCEPTION

The apparent motion of visual targets after cessation of nystagmic eye movements following an angular[20] acceleration (ie spin recovery) or a linear [48]one are termed **the oculogyral & oculogravic** respectively.

Moreover research data on squirrel monkeys[15] suggest that time constant of primary vestibular afferents reflecting cupula[*] long time constant,that is time taken for the cupula to regain normal resting position, is shorter than postrotary nystagmus time constant.

Thus, cupula time constant , as estimated by behavioral methods (*oculogyral effect*) may suggest central neural mechanisms' time constant that is required so that CNS events subside.

Contributory to the above comes the identification of lower threshold values for angular acceleration perception when the oculogyral effect is utilized as a vestibular responce indicator.

In flight liminal aircraft motion can precipitate apparent motion of off-cockpit visual targets without self-motion perception and engage the pilot in a disorientation status.

Oculogravic illusions are also present when there are changes in only the magnitude of the force vector, as in an elevator (**Elevator Illusions** [48])

6. MOTION PERCEPTION LIMITATIONS IN A NON-1 G ENVIRONMENT

There are some fundamental differences between canals & otoliths, which account for their different behaviour in an altered-1G environment:

Structurally, the canals contain a neurosensory mechanism based on endolymph inertia and similarity in density between endolymph and cupula.
Instead, the otoliths rely their proper function on density difference (ratio 3:1) among stratoconia and endolymph.

Functionally,canal afferent activity conveys the differential signal between them, because in the brainstem mutual inhibitory synapses regulate a resting signal of zero.
Otolith afferent activity conveys the summed output of bilateral activity, which does not cancel in the brainstem fully ; sustained activity is essential for the maintenance of postural tone against the sustained 1G.

The above enable canals to act independently of altered gravity, while absence of otolith signal in space induces *Space Motion Sickness Syndrome*.

However, in a zero G environment a person may feel inverted.The upside-down position is the one in which utricla discharge is the lowest [9], and so absence of gravity is misinterpeted as low utricle discharge - **The Inversion Illusion.**
In hyper-G environment, atypical over stimulation demonstrates illusory perception of the G-excess type.

7. VESTIBULAR PERCEPTION LIMITATIONS & MOTION SICKNESS

The vestibular system acts both as a detector of head & body attitude with respect to gravity and as the origin of reflex activity to improve motion control, even when motion is in progress.

Physiological (structural & functional) limitations of the vestibular organs are seen to serve as the origin of non- or mis-perception of motion stimuli and facilitate under certain given flight conditions the onset of illusory

perceptions which constitute disorientation.

On the other hand, a cornerstone fact in motion sickness research is that a functional vestibular system is essential for the occurrence of motion sickness .
Bibliography on studies of labyrinthectomized animals & men converges on their immunity to Motion Sickness.[30,52,54,64]

Therefore, both motion sickness and disorientation originate and utilize the same structural & functional mechanisms, which all bear a vestibular signal basis.

We can accept that the vestibular system provides a motion adapting three dimensional frame with regard to gravitational vertical for all visual, auditory, cutaneous and other cues to be integrated within , so that the image of the surrounding world can be constructed.

However, motion sickness - better described as *motion sickness syndrome* - appears when a sensory neural mismatch signal "poisons" [44] CNS structures responsible for integration of multisensory cues and determining the idea of the outside world and body's position within, according to the neural mismatch theory [25,29,32,36,49,51] .

Vestibular qualities offer a steady basis for most disorientation illusions, which pilot is often unaware of(Type I disorientation - Benson 1988,[2]). Still,there are cases such as the Coriolis Illusions where the pilot is under both disorientating illusions simultaneously with strong nausiogenic feelings (disorientation Type II [2]-where the pilot is aware of his/her disorientation).

Although orientation and motion sickness are physiologically interconnected, their different behavioral characteristics could probably conform with an occupation of different parts of orientation CNS ascending pathways, either in CNS level or in various neural perception sub-circuits involved.

8. DISCUSSION

With time, the concept of Situational Awareness has been realized as being crucial in air mission success & survivability mostly because modern aircraft employ comparable technologies.It is not surprising that pilot performance variations are strongly attributed to the degree of tactical Situational Awareness they possess.

Therefore, elaboration upon modern cockpit technology has developed the concept of situational awareness assistance in three mejor areas:

I.Cockpit Architecture, with the introduction of

(a) Head-Up-Displays(HUD) and their enchancement by angle, symbology design & size [47]

(b) Auto-Recovery aircraft systems, enabling the aircraft to engage an automatic level-off flight pattern when neccessary,

(c) Three-dimentional (3-D) tactical Displays,

(d) Stereoacoustic sounds [45,46], for better use of human resources

(e) Head-Mounted Displays (HMD) and Night Vision Goggles (NVG)

II. In Training, by

(a) Becoming fluent in Instrument Flying Techniques, aborting all external visual cues, and
(b) Receiving Flight Simulator Training with realistic tactic frames.

III. Pilot Selection

With a flourishing research being directed towards the definition of methodologies for situational awareness assessment, such as: Situational Awareness Rating Technique (SART)[57], Situational Assessment Global Assessment Technique (SAGAT-[14]) etc.

Physiological Vestibular Limitations hopefully may not endanger future sophisticated flight safety by means of disorientation illusion mishaps, but will always present a prequisite for all flight information presentation to the pilot and therefore a problem in cockpit integration of modern tactic S.A. aids technology (ie HMD, NVG, etc)

8. REFERENCES

1. BAKER R, PRECHT W & BERTHOLZ A: Synaptic Connections to cochlear motorneurons determined by individual nerve branch stimulation in the cat, Brain Research, 1973, 64, 402-406.

2. BENSON AJ (1988). Spatial Disorientation - General Aspects, & Sp. Disor. Common Illusions. In : Aviation Medicine, by J.Ernsting & PF King, Eds Buttersworth, London.

3. BENSON AJ (1982): The Vestibular System, in Barlow HP & Mollon JL (Eds) The Senses, New York Cambridge Univ. Press.

4. BENSON AJ (1988). Motion Sickness, In : Aviation Medicine, by J.Ernsting & PF King, Eds Buttersworth, London,

5. BENSON AJ, BROWN SF (1989) Visual Display Lowers Detection Threshold of angular, but not Linear, whole-body Motion Stimuli: Av, Sp & Env. Medicine, 1989, 60, 7, 629-633.

6. BENSON AJ, HUTT ECB, BROWN SF : Thresholds for the perception of whole body Angular Movement About a Vertical Axis: Av. Sp. & Env. Medicine, 1989, 205-213.

7. BENSON AJ, KASS JR, VOGEL H (1989): Thresholds of perception of whole body linear oscillation: modification by spaceflight. Exp. Brain Res.

8. BENSON AJ, SPENCER BA, STOTT JRR (1986) Thresholds for the detection of

the direction of whole body, Linear Movement in the horizontal Plane, Av.Sp. & Environmental Medicine,1088-1096.

9.BOFF KR & LINCOLN JE (1988) Engineering Data Compendium : Human Perception & Performance, AAMRL Wright-Patterson AFB ,USA Ch 3.2 (pp 766-788).

10.CARLSTROM D & ENGSTROM H (1955) The ultrastructure of stratoconia Acta Otol. 45 14-18.

11.COLLAR AR (1946) On an aspect of the accident history of aircraft taking off agt night. Reports & memoranda No2277 London : Aeronautical Res.Council.

12.CUTHOYS IS BLANKS RHI & MARKHAM CH (1977) Semicircular canal functional anatomy in cat, guinea pigand man Acta Otol. 83 258-265.

13.DEECKE L, SCHWARZ DWF & FREDRICKSON JM (1977) Vestibular responces in the rhesus monkey ventroposterior thalamus II Vestibulo-proprioceptive convergence at thalamic neurons Exp.Brain Res. 30 219-232.

14.ENDSLEY MR (1990) A methodology for the objective measurement of pilot situational aeareness .AGARD CP 478, pp 1.1-16

15.FERNANDEZ C, COLDBERG JM (1976).Physiology of peripheral neurons innervating otolith organs of the squirrel monkey. III.Responce Dynamics.J.Neurophysiol.39:996-1008.

16.FERNANDEZ C, GOLDBERG JM & ABEND WK (1972)Responce to static tilts of peripheral neurons innervating otolith organs of the squirrel monkey J.of Neuroph., 35,978-996.

17.FREDRICKSON JM & SCHWARZ DWF (1977) Vestibulo-cortical Projection. In Nauton RF (Ed) The Vestibular System. New York Academic Press.

18.GACEK RR (1975) The innervation of the vestibular labyrinth,In Nauton RF (Ed) The Vestibular System. New York Academic Press.

19.GILSON RD, GUEDRY FE, HIXON WC & NIVEN JI (1973) Observations on perceived changes in aorcraft attitude attending head movements made in a 2G bank & turnAerosp. Medicine 44, 90-93.

20.GRAYBIEL A (1952) The oculogravic Illusion Arch.of Opthalm. 48,605-615.

21.GRAYBIEL A, JOHNSON WH, MONEY KE,MALCOLM RE & JENNINGS GL (1979) Oculogravic Illusion in response to straight-ahead acceleration of a CF-104 aircraft. Av.Sp. & Env. Medicine 50, 382-386.

22.GREVEN AJ, OOSTERVELD WJ, REDEMAKERS WJAC (1974).Linear Acceleration Perception.Arch. Otolaryng., 100:453-459.

23.GUEDRY FE (1974) Psychophysics of vestibular sensation, In Kornhuber HH(Ed) Handbook of sensory Physiology (Vol6/2) New York, Springer.

24.GUEDRY FE (1991) :Motion Sickness and its relation to some forms of Spatial Orientation, AGARD LS-175,pp 2.1-2.23, AGARD/NATO,1991.

25.GUEDRY FE (1970).Conflicting sensory orientation cues as a factor in motion sickness.In 4th Symposium on the role of the vestibular organs in space exploration.NASA SP-187: 45-2.Office of Technology Utilization,NASA,Washington DC.

26.GUEDRY FE & BENSON AJ (1978).Coriolis cross-coupling effect :Disorianting, Nausiogenic or not? AvSp & Env Medicine 49 (1): 29-35.

27.HILDING AC (1953): Studies on the otic labyrinth III: On the threshold of minimum perceptible angular acceleration Annals of Ot.Rhin. & Laryng. 62 5-14.

28.HOWARD ID (1986).The Vestibylar System, in Boff KR,Kaufmann L,& Thomas JP(eds) Handbook of perception & Human performance Vol I pp 11.1-30, New York, Wiley.

29.IRWIN JA (1881).The pathology of Seasickness, Lancet, 2:907-909.

30.KENNEDY RS, GRAYBIEL A,MCDONOUGH RG & BECKWITH RD (1968). Symptomatology under storm conditions in the North Atlantic in control subjects and in persons with bilateral labyrinthine defects, Acta Otolar. Stockholm, 66: 533-540.

31.KUBIN L,MAGHERINI PC,MANZONI D & POMPEIANO O (1980) Responses of lataral reticular neurons to sinusoidal stimulation of labyrinth receptors in decerebrate cats J.of Neuroph.44 922-936.

32.LANDSBERG MP (1960).A primer of Space medicine, Elsefveir, Amsterdam.

33.LINDEMAN HH (1973) Anatomy of the otolith organs Adv. in Ot.Rhin.& Lar. 20 405-433.

34.LOE PR, TOMLO DL & WERNER G (1973) The neural signal of angular head position in primary afferent vestibylar nerve axons J. of Phys. 230 29-50.

35. MATSUNAMI K & SATAKE H (1991).Comparison of monkey cardiac responce to long & large centrifugal acceleration before land after labyrinthectomy. Presented to ASMA 62nd annual Scientific Meeting.

36.MCINTYRE AK (1944).Notes on Airsickness, Austr.Flying Personnel Research Commitee Rep. FR 10.

37.MC NAUGHTON G (1984) Vision in Spatial Disorientation and Loss of aircraft attitude in Human Factors Considerations in High Performance Aircraft Symposium, Williamsburg, VA,USA.

38.MALCOLM R (1983) The Malcolm Horizon: History and Future NASA CP2306:Peripheral Vision Horizon Display (PVHD) Edwards AFB,CA 11-40.

39.MALCOLM R(1974) A mechnism by which the hair cells of the inner ear transduse mechanical energy into a modulated train of action potentials. J of Gen.Phys. 63 757-772.

40.MARINI G, PROVINI L & ROSINA A (1976) Gravity responses of Purninje cells in the nodulus. Exp. Brain Res. 24, 311-323.

41. MEIRY JL (1978). The Vestibular System and Human dynamic space Orientation . Report T-65-1 Cambridge, MA:MIT, 1965.

42.MELVILL JONES G (1957) A study of current problems associated with disorientation in man-controlled flight. Fl. Personnel Res. Commitee Rep. No 1006, London Air Ministry.

43.MENU JP & AMALBERTI R (1990): Factors involved in tactical situation awarenessand the future of support systems AGARD CP-478 OV-F 1-11 .

44.MONEY KE (1991). Signs & Symptoms of Motion Sickness and its basic nature.AGARD LS-175,pp 2.1-2.23, AGARD/NATO,1991.

45.NAISH LN (1989) The simulation of localized sound for aircraft cockpits, RAE Technical Memorandum.

46.NAISH LN (1990) The simulation of localized sounds for improved situational awarness AGARD CP 478 pp 12.1-9.

47.NEWMAN RL (1987) The HUD in Spatial Disorientation in 20th European Symposium , Society of Exper. Test Pilots,Linkoping,Sweden.

48.NIVEN JI, WHITWSIDE TCD & GRAYBIEL A(1963) The elevator Illusion: Apparent motion of a visual target during Vertical Acceleration .Fl. Pers. Res. Commitee Rep. No 1213.London Air Ministry.

49.OMAN CM (1940).Motion Sickness: A synthesis and evaluation of the sensory conflict theory Comp.J.Physiol.Pharmacol. 68:294-303.

50.PELLIEUX L,PIEDECOQ B & QUANDIEU D (1989) Saises d' informations par voie auditive Rapport CERMA No 89-29.

51.REASON JT (1978)Motion Sickness Adaptation: a neural mismatch model, J.Royal Soc. Med. 71:819-829.

52.REASON JT & BRANDT JJ (1975).Motion Sickness,Academic Press, London.

53. SANTUCCI G, AMALBERTI R & BARRAULT B (1984) Enquete sur les ddesorientations spatiales sur super etendard Rapport CERMA 84-02

54. SJOBERG A(1929).Experimental Studies og the eliciting mechanism of seasickness,Acta Otolar. 13:343.

55. SMITH C (1956) Microscopic Structure of the Utricle Ann.of Ot.Rhin & Lar. 38 131-138.

56. STOTT JJR (1986) Mechanisms and treatment of Motion Illness , in Nausea & Vomiting: Mechanisms & Treatment C.J.Davis, Lake-Bakaar, and DG Grahame-Smith,Eds.,Springer-Verlag,New York,110-129.

57. TAYLOR RM (1990)ituatonal Awareness Rating Technique(SART) The development of a tool for aircrew systems design. AGARD CP 478, pp 3.1-17.

58. TAYLOR MM & CREELMAN CD (1967) Efficient estimates of probability function J. of Acoustical Soc. of America 41 782-787.

59. TJERNSTROM O (1977) Effects of Middle Ear pressure on the inner ear.Acta Otol. 83. 11-15.

60. TRINCKER DEW (1962) The transformation of mechanical stinulus into nervous excitation by the labyrinthine receptors.In Symp. of the Soc of Exper.Biology No 16, BiologicalReceptor Mechnisms, Cambridge Univ Press.

61. TUMARKIN IA (1937) Some observations of the function of the labyrinth Proceedings of the Royal Soc. of Medicine, 30 599-610.

62. WALSH EG(1961).Role of the vestibular apparatus in the perception of motion on a parallel swing J.Physiol.155:506-513.

63. WALSH EG (1962)The perception of rhythmically repeated linear motion in the horizontal plane Br. J.Psychol. 53:439-445.

64. WHANG SC & CHIN HI (1956)."Experimental Motion Sickness in Dogs: importance of labyrinth and vestibular cerebellum" Am.J.Physiol. 185: 617-623.

65. WILSON VJ, GACEK RR, MAEDA N &UCHINO Y(1977) Saccular and utricular inputy to cat neck motorneurons. J. of Neuroph., 40, 63-73.

66. WULFFTEN PALTHE Van PM (1942).Function of the deeper sensibility and of the vestibular organs in flight. Acta Otolaryngolog. 4 415-448.

67. YOUNG LR (1984).Perception of the body in space: mechanisms IN: handbook of Physiology - The Nervous System III 1023-1066. Washington DC: Am.Phys.Soc ; 1023-60.

AGARD Neuroscience Group Linking Lecture

"Neurological Dimensions of Situational Awareness"

J. L. Firth
Department of Neurosurgery, University Hospital, Queen's Medical Centre, Nottingham NG7 2UH, UK.

The first problem in the Situational Awareness [SA] field is that of definition. The how, why, where and whens? Above all **what is SA?** Neurologically, SA is defined as **the accurate, comprehensive, four-dimensional appreciation of a situation at any one point in time.** To be operationally effective SA has to be continuously up-rated, modified and developed throughout the time course of the operational epoch. The operational epoch is the significant period of time under consideration. This can be the duration of an engagement, a flight sector, a detail, a crisis, an operation, a detachment, a campaign, an electoral session or a historical era. **SA is the first of the three essential, inter-linked and inter-dependent components of aircrew operational ability and the foundation for the other two: appropriate decision and effective action.**

Of what is SA composed? There are five basic components:
1. The detection, recognition and appreciation of the multiple, dynamic, four-dimensional factors, both tactical and strategic of any operational situation [this may or may not be complete].
2. An understanding of the priorities of the situation [which may may not be accurate or the priorities may not be appropriate]
3. An analysis of the significant features of that situation, relevant to the task in hand [which may or may not be correct].
4. Definition of the opportunities and options that exist or present themselves for good or ill during the operation and how they may be exploited, ignored or avoided [this may or may not be realistic].
5. Continuous re-appraisal of the situation as it develops [this may or may not be maintained or uprated frequently enough, whilst the detection of new factors, threats and opportunities may not be achieved].

When is SA important? Throughout the operational epoch.

Where does SA fit in? As the first of the three components of aircrew operational ability-

"To decide and to do
What you have to do,
You first have to know
What's all going on - SA".

Or as they say in St Louis:

"Knowing what the hell's going on to know what the hell to do with it".

ADAM, E[1994]

Why is SA important? Without it the rest of the defence system may be rendered useless or worse than useless.

What does SA do? It provides the basis for apt decision and decisive action [which may be to do nothing].

What are SA's boundaries? All factors and matters appertaining to the situation [in effect potentially boundless].

Upon what does SA depend? The Neurosciences. On human neuroanatomy which is bedevilled by such inconveniencies as a limbic system arranged to suffer maximum disturbance in high G and force fields. Upon human neurophysiology in which 2nd messenger species specificity means that the bottom line is Alexander POPE's aphorism that "the proper study of man, is man". One cannot extrapolate directly from animal models to man. These problems are the subject of the Fall 1995 Symposium which is arranged to follow-on and address the challenges raised this Spring.

What is the neurological substrate for SA? How does it work? Man is a 4-dimensional reflex animal [Fig.1]. The afferent limb of the reflex drives the reflex [and drives us all]. We are children of time and >90% of what is driving us at this moment in time, in The Present ["time-now"], is derived from our past, a

Paper presented at the AGARD AMP Symposium on "Situation Awareness: Limitations and Enhancement in the Aviation Environment", held in Brussels, Belgium from 24-27 April 1995, and published in CP-575.

function of memory. That past may be reality or illusion, innocent or deliberate delusion. For infants a non-stop Television [TV] watching world becomes their real-world. As there are more afferent events per second on TV than in real life, TV-time is that much more educative. What the Media have achieved today is that to which Jesuits, Nazis and Soviets aspired in the past: "to get 'em young". Unfortunately it does not end with adolescence. Target population brain washing was never so effective or so pervasive as it is today.

"Memory" systems operate at every level. Within cells intracellular 2nd messenger modulation reflects past practice and exercise, introducing favoured responses and bias. Between cells, intercellular synapses, transmitters and networks between dendrites, axons and terminals support pathway development and enhancement by repetitive facilitation and disuse atrophy. This leads in time to system specificity and susceptibility moulded by past experience. Systems supported by the central grey reticulum and its analogues; post-natal cerebral hemispheric "colonization" with superabundant pathways; initial over-provision and later developmental decay of under-used pathways [failure of which is reflected in the impairment of megalencephaly]; their serial myelination and then system integration manifest in education as "entry gates"; the phenomena of biological rhythms and reticular hunting; all involve the temporal dimension. Cognition and "conscious memory" based on smell and the limbic system are not the only time-critical mechanisms in human neurological performance. G-field orientation requires four-dimensional critical somatic attitude maintenance which is based on the archicerebellar vermis and its connections, the bane of the agile aircraft operator. Tonic reflex organization depends on the temporally orientated paleocerebellum. Sub-cortical, initially consciously and voluntary acquired, but functionally automatic spinal and limb motor skills are established and maintained by prolonged practice and require the neocerebellum and "motor memory". Bicycle riding is a three dimensional example of these time-critical, sub-conscious but acquired and trained abilities. Flying exploits all four.

To date man has flown subject to the physico-temporal constants under which he evolved. Superagility introduces novel environments dominated by new control laws. For survival in and the effective exploitation of the opportunities of such demanding environments in the short term, Evolution, Natural Selection and the passing of the generations are too slow. Mastery of "time-now", our epoch, requires the succesful resolution of complex real-time tasks in time domains shorter than those in which man and the limbic system developed. To resolve these inevitable incompatibilities and to put them to good use, Man has to team with machine. This is an area of AGARD-AMP's particular responsibility: to ensure that Mk.1 Man is understood in defence design and development, successfully integrated into, optimally employed by and ultimately in command of each system and its deployment.

Faced with these challenges utilization of afferent, sensory physiology relies initially on 3-dimensional touch, temperature, texture, consistency, pain, pressure, position and hunger transduction. With input gating and thresholds, smell, vibration, rhythms, force fields, visual, auditory and spatial stimuli all provide 4-dimensional appreciation of the environment. This is complicated by personal, social and emotional factors and compounded by digital [for which we were not designed] as well as analogue stimuli [for which we were]; image dependency and field specificity; gaze and vision; pattern and target recognition; thalamic thresholds and perception; subliminal automatic activity; sensory integration, diversion, distortion. suppression; inattention and saturation. Together with the problems of preconception, illusion and delusion there is material enough for a series of AGARD symposia.

Reflex responses [segmental, internuncial; across the midline; inter-segmental; supra-segmental and trans-temporal into the fourth dimension, time] to these inputs may be expressed in a reflex, automatic or conscious fashion. All responses are sensitive to pre-signalling, modulation and timing. Efferent, motor output may be reflex, that is congenital or "genetic", in-built and fast, as in insects. Automatic responses, learnt over time by deliberate, repetitive voluntary effort merge with habit and benefit from continued practice. They are slower but require no prior thought within the operational epoch. Once established they operate in the "subliminal", "near-threshold", "second-nature" or "instinctive" domain, traditionally as in "shooting from the hip". Conscious, cognitive action is voluntary, very slow and memory-dependent. It is modulated by pattern recognition [that is, by "experience"], by practice, intelligence, attention, observation and is subject to morale and fatigue.

Central processing, the object of psychological study, obeys the same basic system rules and is dominated by neuro-anatomy, physiology and pharmacology. Man operates as a 4-dimensional physical, spatial and temporal continuum. For effective, efficient, economical defence we have optimise and exploit that continuum to manage or drive the future rather than be ridden by events and the ambitions and activities of others

At the cellular level synaptic, dendritic and axonal transmission introduce their own peculiarities. Hierarchy is the order of the day. Operational hierarchies mimic the neurophysiological. In pairs and teams running from pilot to president all have to possess adequate Situational Awareness to be effective. The vertical integration they represent collapses in crowds. When does a team become a crowd? Crowd awareness, perception and delusion is in its turn confused with herd instinct especially when the comfortable, mindless protection of "safety in numbers" collapses in panic. Much in this important area remains to be learnt from Gideon and Byzantine military experience

Individual interest, preservation of the species and the improvement of society become confused when heirarchy reachs the electorate. Deliberate restriction or distortion of individual and collective SA for tactical and strategic purposes now becomes sensorship, "skewed SA". In the past this was the near-exclusive preserve of Government. It is now dominated by the Media itself, by factional interest, irresponsible parties, aspiring politicians and narrow social and economic groups. Today media manipulation and control is near-universal and so all-pervasive as to be accepted as normal. Wars are lost, but never won by the media. Public information, formerly "news" has been reduced to the media of outrage, the medium craving attention only for itself rather than the public interest. Alternatives, bye-passing independent media to achieve public SA have had only limited success whilst in NATO countries competitive broadcasting and counter-propoganda is only directed abroad. There is a danger that free societies may collapse under the sheer weight of the constant barrage of bias and misinformation. Freedom of or access to fact, to unbiased information may in the future depend on alternative channels of communication. The most obvious candidates at this time are the Internet and the World Wide Web. What started as a secret means of transferring confidential data may yet become the key to the survival of the free, unblinkered and original thought upon which mankind's progress depends.

When is SA developed?

SA development and training starts at conception and is well advanced by the time we leave the womb. Educational "entry gates" and memory dependency make SA time and career sensitive and subject to the phenomena of attention, motivation, practice, familiarity, facilitation and fatigue which involve the whole central nervous system [CNS]. This ensures difficulty when attempts are made to isolate and study individual components of what is functionally a sensori-motor continuum modulated by memory. Handling SA requires a system approach, where the system is the CNS. Hence the formation AMP's multidisciplinary Neurocience Group.

How is SA developed?

Traditionally on the playing fields of Eton. In reality *ex utero semper nova est*. Darwin and 40,000 generations' struggle for survival means that we have an in-bred, potential ability to be "street-wise". But problems of street specificity arise when this is applied to four and more dimensional practice. For some Huntin', Shootin' and Fishin', sporting skills + 5 hours flying training in an Avro 504K may have been adequate preparation for a Sopwith Camel and an afternoon with von Richthofen. But in delivering the goods, the Right Stuff of 1918 and beyond, experience, that is the good fortune to avoid attrition and survive long enough to become an Ace, proved a major factor and was adopted as deliberate policy in the Soviet Air Forces. With present economic constraints such practices are luxuries that NATO budgets cannot afford. Nevertheless it is an essential that we re-create the benefits of operational attrition, if necessary by simulation coupled with the use of cheap, aerobatically unlimited surrogate aircraft following the 1930s Luftwaffe and Red Air Force models.

Can SA be modelled or characterised? It can. The field is beset by a profusion of theoretical models, an *embarrass de richesse*. In considering their relative merits an evolutionary approach is helpful.

To the primordial worm consistency was all-important. Smell required, was and remains memory. Seasons afforded sun-cues, periodicity produced the pineal. Surfacing posed new threats and opportunities and invoked the development of ocular vision. Marine operation favoured the ability to remain upright in a gravitational field, facilitated by the archicerebellum. Efficient, economical swimming required rhythmic, serial control of axial muscle tone by the paleocerebellum. This in its turn made possible a return to terrestialism through the semi-aquatic medium of mud [as in "mud, glorious mud"]. Once there, the attractions of and safety afforded by trees ensured that only those with limbs and the ability to use them survived. Those, that is, who had developed motor memory and the neocerebellum. Trees have nuts. Nuts need tools and parietal lobes to enjoy them. By this time survival in hostile environments North of the Sahara necessitated sharing caves with mothers-in-law. Only those with frontal lobes and capable of empathy came in from the cold. Next, education [smell suborned for "memory"]; the resultant need to delay procreation as "adolescence" and the competitive advantages of training and technical innovation ensured the success of family and tribe. The teams of old were more than the crowds or herds of today. Much learned from our old ally dog has

been forgotten and the objective study of crowd behaviour is neglected in the West. Progress by co-operation in competition and as Society is often hazarded by confusion between individualism and reality.

That being said, for the practical operator a "conveyor belt" or "production line" model of SA is recommended. A minimalist version [Figs.2] contains the fundamental elements to be considered by responsible authorities. For the present debate it illustrates the continuity of situational awareness with decision and action, together with the importance of recognising their differing domains, determinants and priorities. SA should not be confused with Operational Effectiveness, though for many senior NATO staff the two are synonymous. Overall Operational Effectiveness depends on Situational Awareness. They are parts of one process, but they are NOT the same thing. Remember Myriomcephaly. Likewise if aircrews are to concentrate on the developing situation and how to exploit it, they have to be relieved of as much routine business and surveillance as possible. The model, for example, indicates that not having to watch the exit track for anything that is not essential for the exploitation of the tactical and strategic situations as they develop, or better still, are manipulated, will allow greater attention to be paid to them - the whole purpose of the exercise.

Applying this example: effective, reliable voice command and execution of all routine functions that at present require voluntary attention, action and checking [such as radio frequency changes, headings, levels, arming, selecting displays, etc] will avoid diversion of the pilot's attention from the afferent to the efferent limb of the model and allow more time and attention to be paid to the critical input side - the real world of real targets, real threats and real opportunities. To optimise operational effectiveness, the pilot must in this way be able to delegate all routine tasks by simple command [voice being at present the most readily available modality] to subordinate systems, just as a ship's captain does to his lieutenants on the bridge whilst fighting or manoeuvering his ship [*pace* F-15s, Hinds, helicopters and *Vincennes*].

Like a good skipper who has learned his way up through the ranks and understands the abilities and limitations of his craft from bilge to ensign and over and above this has a clear grasp of the operational situation, the pilot has also to gain and retain a Yeageresque ability to run the operation when all else fails, which in commercial practice is where aircrew earn their keep. As with operational effectiveness when all systems are functioning as intended, the pre-requesite of success in the face of systems failure is pre-crisis situational awareness.

Where does the AMP go from here? SA is a major concern and its optimization a priority of civil and military authorities alike. And not only in aviation. It is necessary to define immediate, mid-term and long-term objectives in achieving this and to so order our priorities that AGARD and the AMP can work effectively as a team, addressing the present challenges and taking the opportunity to review progress in Koln in the Fall.

CHECKLIST for AMP SA Symposium

Has the Symposium answered the many basic conserns? The reader should be able to answer 12:

1. What?

1.1. What is SA? An agreed definition?
1.2. What does it do?
1.3. Of what is it composed?
1.4. Upon what does it depend? Upon external, internal, present, past and future factors? Are they functions of piloted, remote, automated, all operations?
1.5. What are its boundaties?
1.6. What are the consequences of failure in this field?
1.7. What concrete benefit will progress provide?

2. When?

2.1. When is SA important?
2.2. When can it lapse or be ignored? Is it always a continuous process?

3. Whence?

3.1. Whence have we come?
3.2. Whence are we going?
3.3. Whence **should** we be going?

4. Where?

4.1. Where does SA fit in?
4.2. Where is it important?
4.3. Where are we now?

5. Whether?

5.1. Whether time should be given to this/other subjects?
5.2. Whether we can ignore the subject?
5.3. Whether we would be better served by "native cunning", by "street-wise" in-dividuals, by natural selection or by operational attrition?
5.4. Whether we should be using other training and enhancement methods [including surrogate unlimited aerobatic aircraft]?

6. Which?
6.1. Which strategies are "best"?
6.2. Which are the key, critical factors?

7. Withall?
7.1. Wherewithall should the matter be pursued?

8. Whither?
8.1. Whither, from whence do we need to draw expertise?

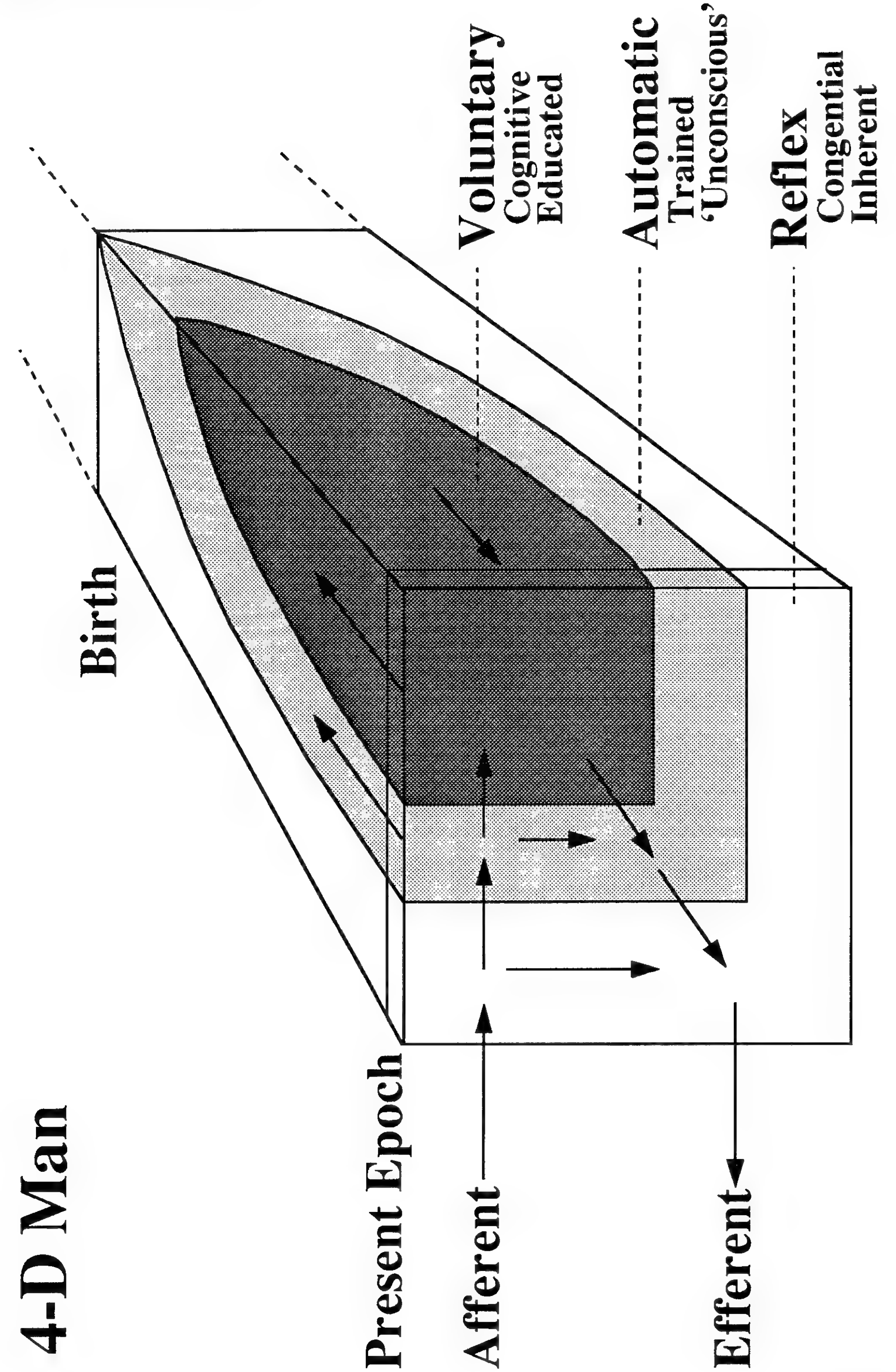
9. Who, to or by whom?
9.1. To whom is SA interesting, important, vital?
9.2. Who is or should be responsible?
9.3. Who should be doing this work?
9.4. By whom should it be co-ordinated?
9.5. By whom should it be funded? On what criteria?

10. Why?
10.1. Why is it important?
10.2. Why particularily at this time?

11. How?
11.1. How does it work?
11.2. How can one educate and select for, train, improve, optimise SA?
11.3. How is SA best monitored?
11.4. How can it be maintained? Cues? Sufficient sleep? Watch systems? Other routines? Pharmacologically?
11.5. Can it be replaced and if so how?
11.6. Do machines do it better? If so how and by how much?
11.7. How do we best integrate man and machine?
11.8. How is situational information best presented to man? In reflex, subliminal, automatic, voluntary, cognitive forms? Are analogue stimuli essential? Can digital cues cope in critical conditions? Is colour vision fast enough?
11.9. How much is SA actually worth?
11.10. How is the Defence appreciation of SA effected by the economic necessity to return to dependency on the ability of small operational units, often individuals, to succeed in critical, potentially irremediable situations [ie. how dependent is the return to Gunboat Diplomacy on SA for its success? On old-fashioned "Local Knowledge" in XIXth century parlance?].

12. Can SA be modelled?
12.1. Available models.
12.2. Model requirements.
12.3. Model design.
12.4. Model selection
12.5. Model validation, monitoring, rejection.
12.6. Model application
12.7. Model development.

Fig. 1 4-D Man


4-D Man
Present Epoch
Afferent
Efferent
Birth
Voluntary
Cognitive
Educated
Automatic
Trained
'Unconscious'
Reflex
Congential
Inherent

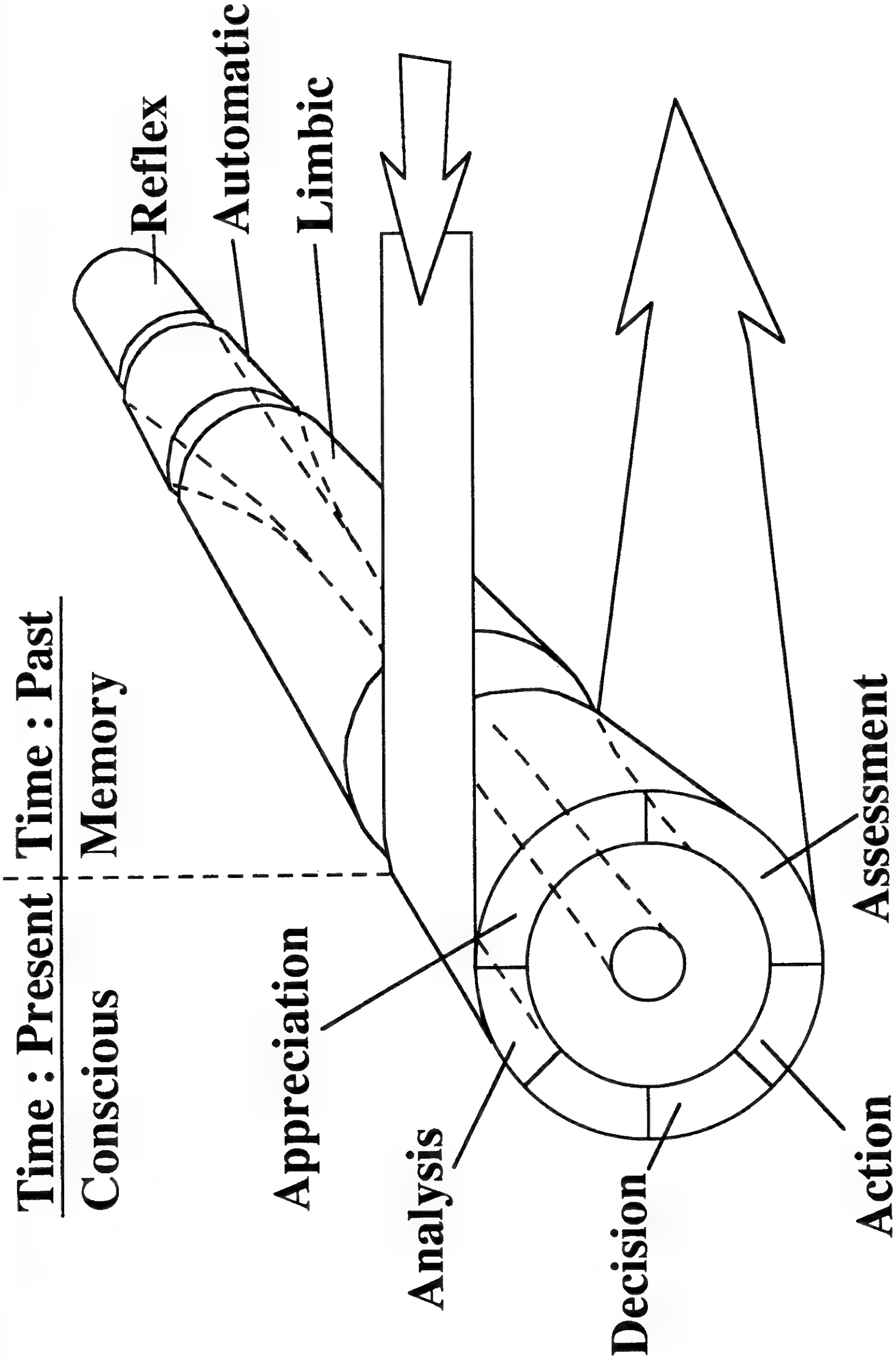
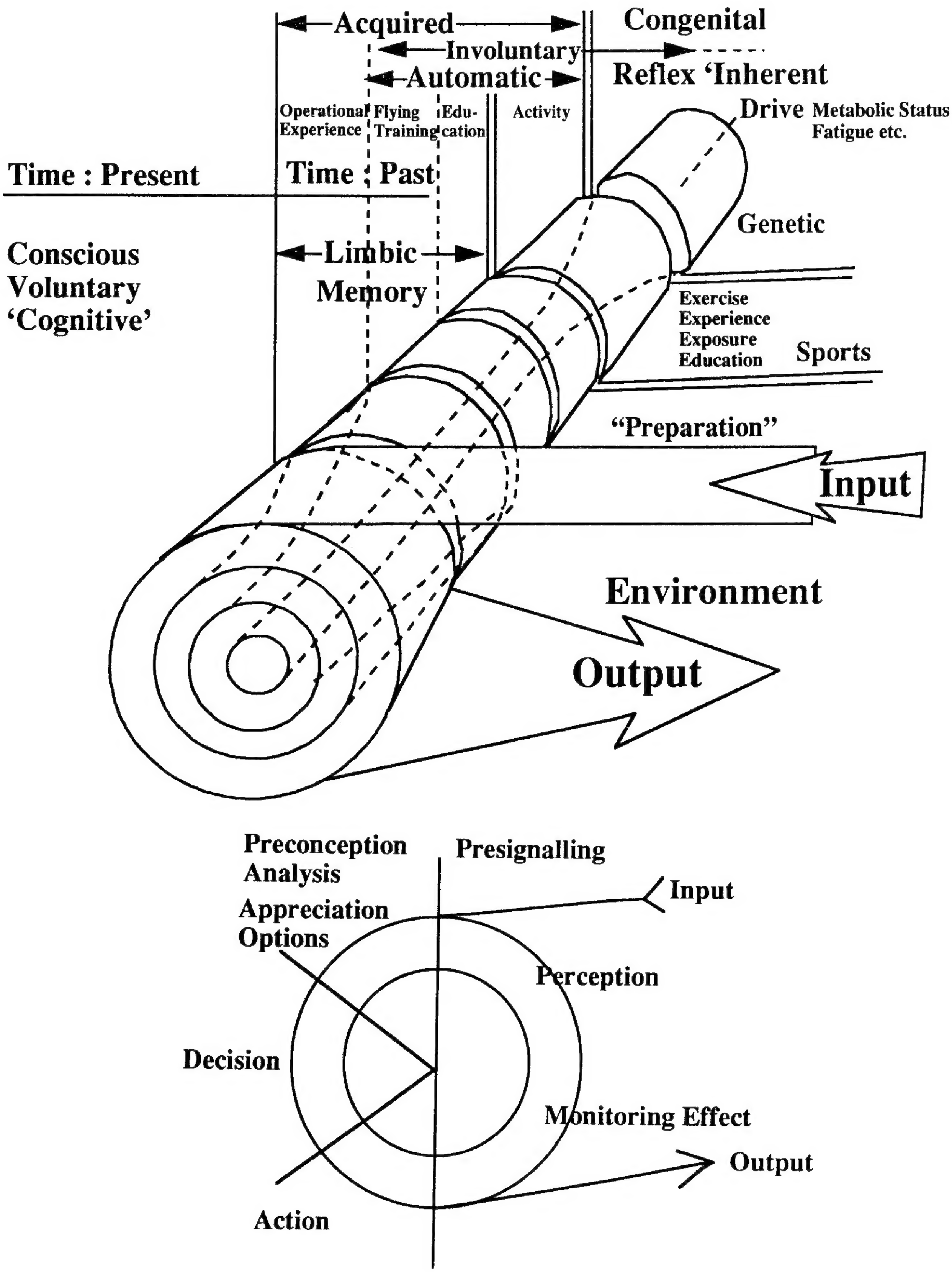


Fig. 2 . SA ‘Conveyer Belt’ Model

Fig. 3. 'Conveyer Belt' or 'Rollingmill' Model of SA



©JLF 1994

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-575	ISBN 92-836-0021-5	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title
Situation Awareness: Limitations and Enhancement
in the Aviation Environment

7. Presented at/sponsored by
The Aerospace Medical Panel Symposium held
in Brussels, Belgium from 24-27 April 1995

8. Author(s)/Editor(s)	9. Date
Multiple	January 1996

10. Author's/Editor's Address	11. Pages
Multiple	268

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Situational awareness
Accident investigations
Pilot training
Workloads
Man machine systems
Aircraft
Cockpits
Cognition
Visual perception
Speech recognition
Laser damage
Display devices
Aerospace medicine
Human factors engineering
Performance evaluation
Man computer interface

14. Abstract

These proceedings include the Technical Evaluation Report, two keynote addresses and 25 papers from the Symposium sponsored by the Aerospace Medical Panel and held in Brussels, Belgium 24-27 April 1995.

Situational Awareness is seen as key to mission success and aircraft safety. There are several questions that the Symposium addressed: how effectively Situation Awareness can be measured, whether it is possible to select for it and whether training strategies can improve it. The Symposium also examined the research carried out into the contribution of new Cockpit Technologies to enhance it. Loss of Situation Awareness has been the predominant cause of fatal accidents in both military and civil aviation and several examples were cited where the aircraft had been lost or put in jeopardy due to pilot error.

These proceedings will be of interest to those involved in cockpit system design, human performance, human perception, cognition and accident investigation.

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie (1)47.38.57.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.* La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD-VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Directeur, Services d'information scientifique
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Pintor Rosales 34
28008 Madrid

ETATS-UNIS
NASA Headquarters
Code JOB 1
Washington, D.C. 20546

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Força Aérea Portuguesa
Centro de Documentação e Informação
Alfragide
2700 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.

D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for
AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
Etats-Unis

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris
France

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax (1)47.38.57.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above. *Please do not telephone*.

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Força Aérea Portuguesa
Centro de Documentação e Informação
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Pintor Rosales 34
28008 Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Headquarters
Code JOB-1
Washington, D.C. 20546

The United States National Distribution Centre does NOT hold stocks of AGARD publications.

Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below. Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for
AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris
France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
United States

Government Reports Announcements and Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)

Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-836-0021-5